The Arbitrage between Service and Infrastructure Competition in the Light of the "Ladder of Investment" Theory

Henten, Anders; Skouby, Knud Erik

Published in:
Infrastructure versus Service-Based Competition: The Case of Mobile Telecommunications

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

A collection of Essays on

Infrastructure versus Service-Based Competition: the Case of Mobile Telecommunications

Edited by

Laurent BENZONI

Patrice GEOFFRON

Foreword by Martin CAVE

QUANTIFICA

QUANTIFICA PUBLISHING
32 RUE DES JEUNEURS
75002 PARIS
FRANCE

ISBN : 978-2-9529662-1-4
EAN : 9782952966214

CONTENTS

Part II : Focusing on Mobile Issues

Foreword

MARTIN CAVE
Director, Centre for Management under Regulation
Warwick Business School

The issue of the relative merits of infrastructure and service competition has been a feature of debate in telecommunications regulation for decades. There are risks for a regulator pursuing either policy.

Infrastructure competitors may prove to be weak and ineffectual, and a failed reliance on this form of competition can consign end users to the mercies of an inadequately regulated incumbent in an environment of high prices and little choice.

Alternatively, by making *service competition* easy and profitable, regulators may deter investment in competing infrastructures which could offer customers a genuine choice among differentiated services.

The mistake here is to neglect the Janus-faced relationship between regulatory intervention and market structure. Regulation must respond to market power. But regulatory interventions also affect investment incentives, which influence market structure. A regulator which underestimates the scope for infrastructure competition is often making a self-fulfilling prophecy.

To some degree, the dilemma was 'solved' by reference to the so-called 'ladder of investment', with which my name is associated. This saw infrastructure competition as a progressive process, in which a competitor could begin by reselling the historic monopolist's service and then gradually move nearer the customer by making its own investments in a core network, in backhaul, and ultimately by installing electronics in the local exchange and relying on the access provider solely for a rented copper loop. In this framework, the regulator would seek to move competitors up the ladder, as they acquired more customers and could make profitable investments in less easily replicated assets.

A ladder is a powerful aspirational metaphor, and many European regulators have adopted the approach. Unfortunately, it is difficult to test a complex dynamic hypothesis such as the ladder of investment, but it is noteworthy that in the countries in the EU in 2003, broadband (DSL) competitors relied predominantly on resale. By 2007, unbundled local loops provided most of their connections and resale was much less important.

But this prospect of a comfortable progression to infrastructure competition has been shaken by new fibre-based next generation access (NGA) technologies which adopt a different architecture than existing networks. In

many cases, the competitor's potential access point at the local loop disappears, leaving access competitors the option of either taking their fibre to the sub-loop which is even closer to the customer (a strategy which may not be technically or commercially viable) or of withdrawing to a wholesale broadband access or bitstream product which requires them to rely more heavily than in the past on the incumbent.

In my view, regulators should not resign themselves to the view that a next generation access network is necessarily a monopoly, and that the best we can aspire to in future is service competition or a very attenuated version of infrastructure competition. To the contrary, regulators should first promote end-to-end competition between fibre networks and the alternative NGA networks represented by upgraded cable companies, and consider opening up passive assets such as ducts to promote competition. The growing potential for broadband delivery of wireless networks, which are inherently naturally competitive) should also be taken into account. If these forms of end-to-end competition are deployed, the need for regulation may even disappear.

In this panorama, this collective volume is a useful contribution as is it focused on the specific case of mobile telecommunications. In the aforementioned debate, mobile markets are indeed singular: since their inception, they have been characterized by facility-based competition, contrary to the fixed broadband market that was frequently based on a monopoly in local access. From that viewpoint, mobile markets raise original empirical and theoretical questions. The issue is not that of "climbing the ladder" to create progressive infrastructure competition (that exists by construction), but that of introducing a degree of service competition without deterring future investment in network infrastructures (that would weaken infrastructure competition).

In addressing specifically the mobile markets in the classical "infrastructure vs service" competition debate, the present volume is both relevant and timely. I commend these interesting contributions to policy makers, to regulators, to those working in the industry, and to those who, like me, study it.

Martin Cave
23 July 2008

Introduction

LAURENT BENZONI
University of Paris II (Panthéon-Assas)

PATRICE GEOFFRON
University of Paris-Dauphine

In network industries, two regulatory means can be chosen to introduce competition:

- Infrastructure-based competition, whereby operators are incited to invest in infrastructure and fully compete on both the infrastructure access market (upstream or wholesale market) and on services delivered through infrastructure (downstream or retail market).
- Service-based competition, whereby service providers have an equal and non discriminatory access to a unique monopolized infrastructure, is considered an essential facility and required to offer services in competition on the retail market.

Infrastructure-based competition is obviously preferred because all monopolies disappear. As a result, the optimal regulation process manages the transition from service-based competition to sustainable infrastructure-based competition.

Both of these regulatory means have been experimented across Europe's telecommunications sector. Indeed, the initial "service based competition" regulatory framework was implemented for fixed communications (voice and data), whereas mobile communications implemented "infrastructure based competition".

The right balance between competition via infrastructures or services is an ongoing and controversial debate among academics, as well as regulators. The exchange of ideas has induced innovative concepts, as the "ladder of investment" developed by Pr. Martin Cave. But, if discussions have led to useful guidelines in selecting relevant regulatory schemes, fixed broadband has been primarily addressed (especially around the "ladder of investment" theory). This book is, therefore, a collection of essays assembled to highlight the specific dimensions of these issues within the field of mobile telecommunications.

Indeed, considering competition in mobile telecommunications through this "lens" is particularly relevant today, for both theoretical and regulatory reasons:

- First of all, the mobile industry is a singular example of pure competition via infrastructure, and has been since the very

beginning with very few other comparable cases, be it in telecoms, energy, railroad, postal services, air transport, …

- This basically means that mobile infrastructures have been duplicated along the entire value chain, without any element considered as an "essential facility", the number of operators only being limited by the scarcity of spectrum (leading to national markets from 2 to 5 players in Europe, for example).
- Whereas the national markets where oligopolies, by construction, the aim to increase the competitive intensity in the retail markets, via the introduction of service competition, arose through the regulatory authorities, conducing to the development of radio-less operators[1].
- And, from this viewpoint, we are here in front of the "uniqueness" of the mobile case, as the ongoing process means introducing service competition in an arena where players compete in infrastructures, contrary to the case of fixed broadband where the matter was to erode the power of monopolies.

These evolutions are at the origin of the present project, as they lead to the following questions:

- Is service-based competition relevant to address the (potential) problems of the mobile retail markets?
- Can the introduction of service competition be efficient to cope with oligopolistic infrastructures as well as monopolistic infrastructures?
- To what extent can introducing service-based competition in the mobile markets not disrupt the deployment of future network infrastructures?

European and non-European scholars have been invited to present their works to be widely circulated amongst regulatory authorities, as well as industrial players and the telecommunications research community. By gathering essays addressing the aforementioned interrogations, we seek to propose academic contributions for future regulatory choices. The aim of this collection is that regulatory choices will be consistent with such academic contributions.

These essays are organised in two parts.

1 *Mobile Virtual Network Operators (MVNOs).*

- ***Part I: Theoretical Framework and Regulatory Schemes***

The first part presents the main economic mechanisms related to the alternative "infrastructure/service", the theoretical basis on which these mechanisms rely and the related regulatory schemes.

M.A.Bergman discusses the trade-offs between competition in the entire value chain that comes from infrastructure-based competition and the economies of scale from access-based competition, as well as how balance is affected by the choice of a specific regulatory model. He demonstrates that while Infrastructure-based competition potentially offers less regulation-induced inefficiencies, service-based competition allows the industry to realize greater returns to scale (including network effects). He reviews a number of regulatory options, such as vertical separation, government ownership and access regulation, and analyses their implications in terms of desirability of competition on infrastructure.

A.Henten and **K.E.Skouby** discuss the theory of 'the ladder of investment' in broadband access. This theory states that the entry of players on the market occurs on a step-by-step basis ('rung-wise'). Players enter the market in areas where barriers to entry are low and, later in the process, enter areas with higher barriers. A frame-setting discussion on both the theory and the environment in which it has developed is proposed, highlighted via three country cases (the UK, the US, and Denmark) and emphasises how these countries approached the facility-based vs. service-based competition issue.

J.Kittl, M.Lundborg and **E.O.Ruhle** propose a panorama of the implementation of unbundling by National Regulatory Agencies across Europe and the related impact in terms of economic welfare. This study points out the importance of a balanced approach: it seems that there is no single way towards competition, but that markets need a healthy mixture of both service-based and infrastructure-based competition. They also show that, with convergence and new technologies like broadband and VoIP, access at the levels of networks, services, applications and devices and own infrastructure and thus LLU will become increasingly important for operators to be able to differentiate their products.

U.Blum, **N.Krap** and **C.Growitsch** starting point is the announcement, in summer 2005, of Deutsche Telekom to precondition a 3 billion euro investment in a fibre optic network not to be regulated with respect to pricing and third party access. Unless the new technologically-leading infrastructure was exempted from regulation, Telekom threatened the investment to be made in other areas or countries. On that basis, to develop a regulator's strategy that allows investments to occur but prevents monopolistic prices, the authors model an investor's decision problem under threat of regulation and show that the mere threat of a regulator's intervention may prevent

supernormal profits without actual price regulation. They demonstrate that the regulator can influence both the investment decision and the investor's price via signals on regulation probability and that the concept of "regulating by threat of intervention" is not only applicable to existing infrastructure, but also to new investments as well.

Finally, **K.Hori** and **K.Mizuno** compare service-based and infrastructure-based competition by focusing on a firm's incentive to invest in network infrastructure. They show that when monopoly rent is large, infrastructure-based competition means the initial introduction of infrastructure is undertaken earlier than under service-based competition. However, when both monopoly rent and the degree of uncertainty are small, service-based competition brings about the earlier initial introduction of infrastructure than under infrastructure-based competition. The paper includes discussion of the policy implications of these findings.

- ***Part II: Focusing on Mobile Issues***

The second part proposes various analyses centred on mobile telecommunications and addressing their singularities as regards the infrastructure/service debate.

M.Bourreau, J.Drouard and **L.Ferrali** examine the relevancy of the "ladder of investment" for the mobile industry, and discuss whether the concept could be applied to encourage the development of alternative mobile facilities. They show that the ladder of investment for the fixed broadband market constitutes a regulatory response to a different situation than in the mobile market: unlike the fixed broadband market that was characterised by a monopoly on the local infrastructure, in the mobile market, facility-based competition has been intense since market inception. Therefore, their main conclusion is that social desirability of a ladder of investment in the mobile market does not appear as clear-cut as for the fixed broadband market.

J.P.Fuhr and **S.Pociask** challenge the conventional wisdom that the U.S. wireless market somehow lags behind the European wireless market, concerns that have fuelled a debate over the need to correct the U.S. wireless market, including proposals to add regulations, mimicking a European-style model. Based on data from the OECD and FCC, their analyses show that, in fact, European wireless markets have higher concentration, higher prices and lower usage. Thus the contention that the U.S. wireless market lags its international counterparts is not supported by economic data and no further regulation is needed to address this issue. The authors conclude that a higher degree of facility-based competition may be one key explanation of these to the market's success, as increased facility-based competition in the U.S. market has made wireless providers very

responsive to changes in consumer preferences and led to continual improvement in service quality.

L.Benzoni and **P.Geoffron** propose a comparative analysis of structures (in terms of number of operators) and performances (price, volume, penetration) in the European mobile markets in order to reveal an "optimal structure". It has been shown that markets with three players are those that have the most success in relation to these criteria (or, at least, that markets with four and five operators does not present obvious advantages for consumers). Without concluding that three is a "golden number", these results suggest to authorities that there is not necessarily an economic rationale to add more operators (MNOs as well as MVNOs) that will have to face the persistent first movers advantages of the incumbents. It also indicates that granting an additional license does not lead, in any circumstance, to an increase in the consumers' surplus.

C.H.Vuong discusses the reasoning favouring infrastructure over service-based competition in the context of the European mobile telephony sector. Using principally basic economics and finance toolkits, the author argues that, because of the characteristics of mobile industry (a fast changing industry with high initial irreversible investments) and the complex nature of competition in the European mobile industry, infrastructure-based competition presents more "guarantees" as regards innovation dynamics over the long term. He shows that there is no need to intervene in relationships between MNOs and MVNOs and credible regulatory threats should be sufficient for optimal market outcome, and avoid regulatory costs.

S.Ansari and **R.Garud** analyse the dynamics of the mobile "ecosystem" in the presence of successive technological generations, especially in the case of the 2G/3G "transition". They find that different parts of the ecosystem evolved at different rates with "collateral technologies" influencing the transition path that unfolded. They suggest that, rather than a distinct or unitary shift from an old to a new technology, transitions proceed in a zigzag manner resulting in the emergence of hybrid technologies. Gaining a deeper understanding of the processes and dynamics whereby transitions between generations occur holds important implications for policy and strategy, as to avoid, for the future, the uncertainties that surrounded the path from 2G to 3G.

S.Ulset describes and analyses critical conditions for achieving net benefit from opening the value-chain in mobile communications by introducing mobile virtual network operators (MVNOs). The author reminds us that preliminary experience from the early development phase suggested that MVNOs offering complex bundles of innovative value-added services would not be competitively sustainable as separate firms. The main reason was that achieving both higher revenue and lower transaction costs under simple market contracts proved difficult, however, since increasing sales of

innovative value-added services did not only increase revenue, but also required more complex and less standardised interfaces that increased transaction costs. Even if this "equation" is still relevant (more sophisticated services mean more complexity in contracts), the author indicates that the MVNO market recently entered into more mature phases, supported by more entry-friendly regulatory, technical and contractual practices, increasing the sustainability of the virtual operators.

Finally, **F.Maiorano** and **J.Stern** study the relationship between regulation and performance in the mobile telecommunications sector, taking account of the economic impact of telecommunications infrastructure on aggregate income and of the role of country institutions in promoting economic growth. They address these questions by estimating a system of equations for a panel of 30 low and middle-income countries over the 1990 - 2004 period. The evidence they present confirms the positive effect of regulatory institutions on telecommunications penetration and also highlights the contribution of a more widespread mobile telecommunications infrastructure to higher levels of GDP per capita.

*****

At the end of this introduction, the editors may give their own precise views on the issues addressed in this collective book[2]. This view in consistent with the conclusions of the previous book presented in the same collection and entitled «*Competition and Regulation with Asymmetries in Mobile Markets*»[3].

The works included in this book analysed the situation of asymmetries in the mobile markets due to staggered assignment of licenses and the related structural imbalance between first movers and later challengers because of the advantages and/or behaviours of the actors in place (brand-loyalty or recognition, club effects, …). And this is, from our viewpoint, in the same context that the appropriate degree of service competition in mobile markets shall be analysed and determined:

- Since, at least in Europe, the infrastructure-based competition in biased by these persistent first movers advantages, it must be proven that more service-based competition would really mean more welfare instead of more fragilisation of the smaller MNOs that were, mostly, the later entrants.

2 *View that is not necessarily shared by all the authors, as the diversity was the rule in inviting scholars to join the present project.*

3 *Benzoni L., Geoffron P. (eds.) (2007), "Competition and Regulation with Asymmetries in Mobile Markets", Quantifica, 2007.*

- In other words, regulatory authorities shall carefully analyse the risks of paradoxical effects of service competition in mobile markets, as it may limit the capacity of the smaller network operators to still really challenge the incumbents, then reinforcing the later ones and restricting the intensity of infrastructure competition.

Thus, the risk is present that more “service-based competition” would mean less “infrastructure-based competition” in mobile markets, with a regulatory answer to this threat that is not trivial. We hope that the studies presented in this book may shed light on future regulatory decisions and help to “fine tune” the regulation, so as to preserve the cycles of investment in the mobile networks infrastructures.

Laurent Benzoni *Patrice Geoffron*

Part I
Theoretical Framework
and
Regulatory Schemes

Competition in Services or Infrastructure-based Competition? A Review of the Regulatory Schemes[4]

MATS A. BERGMAN
Uppsala University and Södertörn University College

***Abstract**: It is often proposed that competition in infrastructure is better than access-based competition. This article argues that the benefits from competition in infrastructure must be balanced against the returns to scale – including network effects – that an access-based regime permits. It discusses returns to scale, regulatory costs and benefits from competition in telecom markets. A number of regulatory options, such as vertical separation, government ownership and access regulation, are introduced and their implications for desirability of competition in infrastructure is analysed. Some more advanced regulatory schemes that have been proposed specifically for the telecom sector are also briefly mentioned.*

INTRODUCTION

An often repeated mantra among those who are concerned with telecom regulation is that ex-ante regulation should give way to ex-post regulation. In other words, telecom-specific regulation will eventually have done its work and general commercial legislation, including competition law, will suffice to ensure vigorous competition. "Where telecoms markets tend towards effective competition, we no longer need sector-specific regulation", as the responsible EU commissioner said when the telecom reform package was released on November 13, 2007[5].

This "sunset vision" of telecom regulation is based, to a large extent, on the view that competition in infrastructure is both possible and preferable to

4 *This article is based on a report commissioned from the author by the Swedish National Post and Telecom Agency (PTS) for a conference in 2004. The report was subsequently published in "An anthology on the foundations for competition and development in electronic communications markets", PTS, 2005, available at http://www.pts.se/Dokument/dokument.asp?ItemID=4721.*
A revised version is available at www.sh.se/matsbergman and www.nek.uu.se/faculty/bergman/index.html.
All opinions expressed in this article are my own and do not necessarily reflect those of any institution to which I am or have been associated.

5 *Commissioner Viviane Reding in the EU Commission's press release IP/07/1678 on November 13, 2007.*

access-based competition. It is possible, according to this view, because new technology has reduced scale economies and it is preferable because regulation distorts investment incentives. The theme of this article is to discuss the trade-offs between competition in the whole value-chain that comes from infrastructure-based competition and the economies of scale that comes from access-based competition – and how this trade-off is affected by the choice of a specific regulatory model. I will question the received wisdom that competition in infrastructure is always preferable.

However, for the trade-off to exist, there must exist some kind of bottleneck in the industry; some critical stage in the production process where there are large returns to scale. In the telecom industry, the bottleneck is some part of the physical network, such as the local loop or the ducts at the periphery of the network, or else the network effects that comes from being able to make calls to a large number of subscribers on the same network or on interconnected networks. An operator that does not have access to the bottleneck on reasonable terms will not be able to compete effectively – or not at all.

For the trade-off to exist, there must also be costs associated with regulation. Otherwise there would be no reason *not* to rely on access regulation. An ideal regulation in a market with perfect information and well-informed and benevolent regulators would certainly be able to perform as well as a perfectly competitive market, or even better. In practice, however, regulation will always give rise to inefficiencies. Since the conditions for the ideal regulation will never be satisfied, regulation will always come at a cost, no matter what regulatory model is chosen. Depending on how well an unregulated market would function and on how efficient regulation would be, the best course of action may either be to accept a certain degree of inefficiency in an unregulated market – or to introduce regulation.

Similarly, (unregulated) perfect competition requires ideal circumstances, such as perfect information and price taking by every firm and every consumer. At best, some markets can be approximately perfectly competitive. In practice, most markets, including many telecom markets, deviate quite substantially from the perfectly competitive ideal. The presence of bottlenecks, for example, will give market power to the firm that controls the bottleneck. Market power, in turn, creates distortions and inefficiencies in the market that are costly and that hurt consumers. In telecom, policy-makers have opted to introduce regulation to minimize those inefficiencies, while holding out the unregulated market as the ideal.

This brings us back to the cost of regulation. There are many ways to make regulatory mistakes, some more costly than others. The regulator's situation is asymmetric in an unrewarding way: it is easy to make mistakes that will be immensely costly, as exemplified by the crisis in the Californian electricity market in 2000-2001, while it is difficult to make improvements that will have

even relatively modest payoffs. All regulatory model have disadvantages. Therefore, sometimes the best course of action is to accept a certain degree of inefficiency and make do without regulation.

THE SOURCES OF MARKET POWER AND THE BOTTLENECK PROBLEM

Since regulation comes at a cost, there must be a reason to introduce regulation. In telecom markets, the fundamental reason is that otherwise some firms would be able to exert market power. In other markets, regulation may be introduced because of negative external effects (environmental concerns, for example) or because of asymmetric information between consumers and producers.

There are two main sources of market power in the industry, both of which are related to returns to scale. First, there are traditional, cost-side returns to scale. Telecommunication is dependent on infrastructure and some of this is so costly that it cannot reasonably be duplicated. Second, there are demand-side returns to scale. These are more commonly known as network effects.

➢ *Returns to scale and returns to density*

In network industries, it makes sense to distinguish between returns to density and other types of scale economies. Returns to density exists when it is more economical for a single firm (or network) to serve a certain segment, area or route. For example, if a single mobile telephony network has the capacity to serve all customers in a rural area, it will inevitably be more costly to set up two networks that cover the same area. Similarly, costs are likely to be larger if two or more networks provide cable-TV or fixed-telephony services in the same area, compared to a single-network configuration.

However, the existence of economies of density does not necessarily imply that (other) economies of scale are large. Absent economies of scale, each operator could be the monopoly provider in a relatively small area, such as a city, and still be fully efficiency. With scale economies, each operator has to serve a large area (a large number of customers) in order to be efficient. With scale economies but no economies of density, the customers of all operators could be spread out over the same total area and the operators would still be efficient. In other words, it would not be inefficient to build overlapping networks.

There exist relatively few systematic studies of scale economies in the literature. A simple way to test for scale economies would be to investigate

whether mobile operators in small countries have higher costs, while the extent of the returns to density could perhaps be investigated by comparing costs in densely and sparsely populated countries. The available evidence suggests that returns to density are high in the provision of infrastructure for fixed telephony and for mobile telephony in rural areas, while not particularly high for mobile telephony in large cities. The evidence also suggests that returns to scale are small or possibly even non-existing at relevant scales.

A final point on returns to scale from the cost side is that the importance of these can diminish not only because of technological developments, but also because of demand growth. Mobile telephony provides a good example. In rural areas, where demand is low, mobile telephony is a natural monopoly. In high-density urban areas, mobile telephony is probably not a natural monopoly. But the technology is the same in both types of regions! It is just that there is so much more demand in the cities. To see this, think about the situation when demand for mobile telephony services was low also in city areas, because so few individuals had handsets.

- ***Network effects***

In a market without network effects, the consumer cares only about his or her own level of consumption (and, of course, for the price, the quality of the product et cetera). In a market with network effects, the consumer cares - directly or indirectly - also for other consumers' levels of consumption. The benefit each person derives from his or her consumption increases as the number of other consumers increases, i.e., with the size of the market or with the number of consumers that are connected to the same network.

In the simplest setting, the number of other consumers of the same product has a direct effect on the (marginal) utility of consuming a unit of the product. An example would be telephones or faxes: a given consumer's utility from having a phone or a fax increases with the number of other consumers that also have phones and faxes, respectively. This type of network effect can be called “one-sided”.

Sometimes consumers may only benefit from other consumers connected to the *same* network. This would be the case in a telephony market where off-net calls are impossible. Another example would be products with incompatible standards, for example personal computers that cannot exchange files with computers of another brand. It is also possible that a consumer derives some benefit from consumers on other networks, but less than if the consumers were on the same network. This would be the case if, for example, it is more costly to make off-net than on-net calls, or if the quality of the former is lower (for example, if interruptions are more frequent).

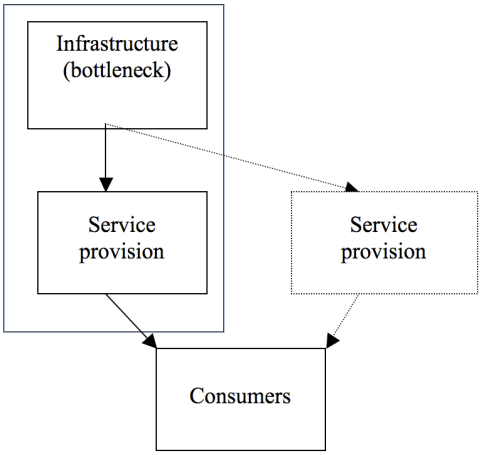
A somewhat more complex situation arises when there are two types of agents that interact on one "platform". Either type cares for the number of agents of the *other* type that uses the platform, but not (directly) about the number of agents of *its own* type that does so. Some examples are buyers and sellers in advertising markets and marketplaces for trading (e.g., stock markets), as well as matchmaking markets (such as dating agencies, real estate agents and business-to-business websites). The internet has some features of two-sided markets: most users of interest are mainly interested in accessing web pages, while some (firms and organizations, mostly) are mainly interested in making web pages available to the general public. Sometimes it makes sense to view callers as one type of agent and call receivers as the other type, even though all individuals sometimes make calls and sometimes receive calls.

A new entrant into a telecom market that cannot interconnect with existing operators would face almost insurmountable entry barriers, exactly because its customers would not be able to benefit from network effects in this market. It follows that the right to terminate calls at a reasonable cost is essential for an operator's ability to compete.

- ***The bottleneck problem***

The telecom industry has a production structure with several vertically related production stages. Because economies of scale and scope vary between the stages, competition is more viable in some stages than in others. This is often illustrated as in the next figure, where only one firm can be active in the upstream infrastructure market, "the bottleneck", while several firms can be active in the downstream market for service provision. For example, the upstream market can be establishing and maintaining a local telecom network (the local loop). The downstream market would then be the retail telecom services market.

Figure 1. The bottleneck problem



Because of returns to scale from the cost side or from the demand side, there will be market power in the bottleneck stage[6]. By itself, this market power gives rise to a number of negative consequences, such as higher prices for consumers and too little production. However, control of the bottleneck can give rise to market power also in the potentially competitive downstream market. In many instances, turnover in the bottleneck stage is relatively small, compared to total turnover. For example, a third of total revenues from fixed telephony in Sweden come from fixed fees (subscription fees)[7].

According to the so-called law of one profit and given that all other production stages are competitive, a company that controls a key stage in the vertical supply chain holds as much market power and can extract as much profit as a firm that has monopolized all production stages of the whole

6 *Returns to scale on the demand side are mainly related to the number of customers already connected to the operator, while returns to scale on the cost side are due to the cost of the physical infrastructure.*

7 *PTS (2003). Arguably, at least some of the interconnection fees could also be seen as payments for the bottleneck controlled by the incumbent operator, TeliaSonera. For comparison, airport costs only constitute about a tenth of the total costs for air travel and the turnover of the central payment systems (e.g., the central giro and card transaction systems) is only a small fraction of total turnover in the retail banking market. Excluding taxes, the distribution and transmission costs account for more than half of the production costs for the electricity bought by a typical household, while this share is lower for large industrial consumers. See Bergman (2002).*

industry. Depending on one's point of reference, this is a good thing or a bad thing. From the consumer's point of view, not much can be gained by introducing competition in the downstream stages (service provision), as long as the network itself is monopolized and as long as the market power of the owner of the bottleneck is not constrained. On the other hand, it is better to have a single-stage monopoly than to have a successive chain of monopolies and highly concentrated oligopolies, since the so-called double-marginalization effect would result in prices even higher than the monopoly price.

REGULATORY REGIMES FOR DEALING WITH BOTTLENECK PROBLEMS

A number of methods have been used in industries with bottleneck problems ("natural monopolies"), all with their respective pros and cons, including the following[8]:

- *Unregulated monopoly*. If competition from substitutes outside of the markets is strong or if demand for other reasons is very elastic, if returns to scale are large and if regulation is likely to be costly, then it may be a reasonable alternative to let a natural monopoly to be unregulated.
- *Regulated monopoly*. If an unregulated monopoly is likely to result in an inefficient outcome (e.g., high prices), if regulation can be relatively efficient and if returns to scale are large – so that duplication is costly – then (consumer-price) regulation of a monopoly provider may be an alternative.
- *Government ownership*. A government-owned monopoly may be an alternative to a privately owned regulated monopoly, in particular if regulation is likely to be inefficient.
- *Franchise bidding*[9]. If an unregulated monopoly is again likely to result in inefficiency (high prices) and if returns to scale are large, then franchise bidding may be an alternative to government ownership or a private monopoly under traditional regulation. In some situations, franchise bidding is informationally less demanding than regulation, since the price is set in a competitive bidding process, rather than by a regulator. An important disadvantage,

8 *Other alternatives are horizontal separation and infrastructural clubs; see Bergman, 2008, for a more extensive discussion.*

9 *I.e., the government auctions the right to serve as the monopoly provider. Payments could go either from the government to the monopoly provider, if the market is unprofitable, or in the opposite direction, if the market is profitable.*

however, is that the bidding for the franchise must be repeated regularly and that the limited franchise tenure will weaken the incentives for investments.

- *Vertical separation*. Sometimes one stage of production where the returns to scale are particularly large can be singled out. If this is the case, and if at the same time an unregulated monopoly would result in inefficiencies, vertically (ownership) separation of that production stage may be a good alternative. The infrastructural stage can either be government owned – or privately owned and regulated. In fact, vertical separation may be useful to limit the negative consequences of regulation or government ownership.
- *Infrastructural access*. If one stage of production has large returns to scale, but if there are also large vertical synergies, then an alternative to vertical separation is to require the firm that controls that production stage to provide access to its rivals. The drawback is that regulation is likely to be more costly than under vertical separation.

Since all of the methods listed in this section have drawbacks, it might seem attractive to try to come as close as possible to the ideal competitive market. That means competition in infrastructure (and services). According to the sunset vision, infrastructure-based competition is the best long-term solution for the telecom industry. If returns to scale are relatively small in all production stages, compared to the costs of regulation, it is easy to agree. However, if returns to scale are substantial, matters are more complicated.

REGULATION AND THE TELECOM MARKET

The telecom market is, and has for a long time been, quite extensively regulated, although the *nature* of the regulation has changed dramatically. Government ownership of national telecom monopolies was the predominant regime in Western Europe before the 1990s, while in the US consumer prices were regulated. In recent years, the trend in most industrialized countries has been towards *access* price regulation. In some respect, this so-called deregulation has substantially *increased* the apparent quantity of regulation as measured, for example, by the number of paragraphs of legal text. However, the telecom firms' latitude to make business decisions has increased, so there is some justification for using the term "deregulation".

Recently, we have also seen an intensification of competition in almost all telecom markets, driven by the entry of a large number of new firms and by a technological development that has lowered entry barriers and created new services that compete with the established ones. Against this background, it is perhaps natural to view the market for telecommunications as developing

into an “ordinary” market, where general competition rules (and other general legislation) will be enough to maintain competition.

In several areas of the telecommunication market, competition in infrastructure has indeed developed. Due to technological progress, long-distance connection, that used to be a natural monopoly, is now competitively provided. The evolution of mobile telephony has resulted in competing networks and there is even some competition for fixed access, in particular in business districts, but also more generally for broadband access. What does this tell us about the need for regulation?

If the technological progress in combination with demand growth lead to a situation where returns to scale are no longer big enough to confer significant market power to the largest firms, then it may in fact be justified to repeal the sector-specific regulation.

Is the conclusion, then, that infrastructure-based competition is both feasible and preferable to access-based competition? I argue that the conclusion is premature, since there still exists infrastructure that it would be inefficient to duplicate. One or two sets of assets have enough capacity to serve the whole market, for example in fixed broadband services. This suggests that policy makers *will* face a trade-off between competition and returns to scale.

In addition, because of the special nature of the telecom market, sector-specific regulation may be necessary even if there are multiple infrastructures. For example, under the 2003 EU Electronic Communication directive as well as under the revised recommendations for which markets that should be subject to ex ante regulation[10], telecom operators are required to provide termination access at cost-based prices, no matter the number of competing networks. This has more to do with demand-side returns to scale (network effects) than the traditional supply-side returns to scale, but it still means that regulation is useful.

10 *November 13, 2007, available at: http://ec.europa.eu/information_society/policy/ecomm/doc/library/proposals/rec_markets_en.pdf*

THE DYNAMIC EVOLUTION OF THE TELECOM INDUSTRY AND ITS REGULATION

While some telecom services are still characterized by returns to scale from the cost side or from the demand side, it is also true that the industry has become much more competitive over the years and that many infrastructural services are now competitively provided. Two prominent examples are mobile telephony network and long-distance voice and data traffic.

When an industry such as the telecom industry is de-monopolised, competition will develop at different speeds in different segments. In most instances, the dynamic development of a market is best handled by the market itself. However, since the development in a bottleneck industry is dependent on the regulatory framework, the policymaker cannot completely sidestep the issue of where competition should first be introduced.

For at least two reasons, it seems natural to introduce competition in service provision before competition is introduced in the provision of infrastructure. Typically, returns to scale (or density) will be larger in the provision of infrastructure than in service provision. Consequently, duplication will be more expensive in infrastructure and, relative to its costs, the returns to introducing competition will be smaller. Furthermore, entry into services market will typically be associated with smaller sunk costs. This means that if one or several competitors enter the market, but competition turns out not to be viable, much less will be lost than had they entered with their own infrastructure.

The “sunset vision” presumes that competition will develop – and regulation will be dismantled – successively. From the start as a regulated or state-owned monopoly, the first stage will be entry into services markets. Because of the bottleneck problem, the entrants need to be assisted by the introduction of access regulation. In the second stage, the entrants will begin building their own infrastructure, but the market structure will still be asymmetric (one firm will be dominant). In this stage, there must still be regulation, but it need not be as stringent as before. In the third stage, the market has become more symmetric and the industry is no longer dominated by a single firm. In this stage, sector-specific regulation can – or so it is presumed – be dismantled[11].

11 *This is perhaps a “telecom-centric” view of regulation. Entry into the tv market has typically been associated with the development of competing infrastructure, while nobody expects entrants into the markets for electricity or railway traffic to build their own infrastructure.*

Full-blown facilities-based competition, the third and final stage, potentially has the advantage that no regulation is needed and, consequently, that the otherwise inevitable costs of regulation can be avoided. These costs include regulatory capture (the risk that the regulators favour the industry, rather than consumers or welfare), regulatory risk (the risk that the regulators over-emphasize short-run competition and therefore reduces access prices so that investment costs cannot be recovered) and bureaucracy costs.

However, there are at least four problems with the above view. The first and most obvious problem is that when there are substantial returns to scale, facilities-based competition means wasteful duplication. Sometimes, technological development or demand growth leads to a situation where duplication is feasible or perhaps even necessary. Sometimes the extra cost of duplication is worth incurring, because the benefits of competition throughout the whole "value chain" are so large. But sometimes economies of scale are large enough for them to be the primary concern, even if that means that one has to live with a less-than-perfect regulation of the monopoly bottleneck. This is likely to be the case for some infrastructural services, such as the local loop or fibre-to-the-household, for the foreseeable future.

The second problem is that in industries where interconnection is essential, such as the telecom industry, it is not necessarily true that regulation can be dismantled when returns to scale from the cost side are no longer important. In order to realise network benefits, subscribers must be able to make off-net calls. If the infrastructure is owned by two or more firms, this means that there must be two-way access – and probably also two-way access regulation. In markets with relatively symmetric firms, two-way access will tend to arise spontaneously, but there is a risk that the access agreements will be anticompetitive, so that they create incentives for high retail prices. In less symmetric markets, large firms may have incentives to foreclose – i.e., not interconnect with – smaller rivals. It should also be noted that interconnection for termination is more problematic than interconnection for origination. There is direct competition between firms that offer origination – i.e., phone services. The customer can chose between operators on the basis of price and quality of service. On the other hand, the active (calling and paying) party cannot chose on which network a given call should be terminated. Unless the receiving party cares for the welfare of the calling party – or fears receiving few calls – there is no incentive for the receiving party to be concerned with the termination fees. This means that competition will not be effective in reducing termination charges. The asymmetry between interconnection for termination and interconnection for origination is reflected in the E-com directive, where the obligations laid on an operator are independent of the operator's market share in the termination market, but proportionate to the origination market shares.

The third problem with the "sunset proposal" is that the process towards balanced facilities-based competition is not an automatic one. Clearly, the choice of access regimes in the first two stages referred to above will influence the development of competitive infrastructure. If the (one-way) access regime is "stringent" (favourable for the entrants), they will have little incentives to build their own infrastructure. If, on the other hand, the access regime is not stringent, there may be no entry at all. Possibly, a well-balanced regulation will result in a situation where facilities-based competition develops over time, but it is likely that for this to happen, the regulator must make active decisions. One example would be access prices that rise over time. Then, it would initially be advantageous to enter the services market. With time, however, it will become more and more advantageous also to invest in infrastructure and less and less advantageous to be active in the services markets only. Obviously, this kind of active policies requires the policy maker to be informed about the desirability of duplication, which in turn requires explicit or implicit estimates of scale economies, of the benefits of competition and of regulatory costs. A policy that that leans too much in favour of competition in infrastructure may result in too much duplication or, alternatively, in too little entry and too high prices. Conversely, a policy that unduly favours competition in services may result in too little facilities-based entry and too little investment by the incumbent.

The fourth problem with the "sunset proposal" is that free-entry facilities-based competition may conflict with concerns for universal service. The new entrants will focus on low-cost high-demand customers, such as densely populated areas and big commercial customers. The incumbent will then be left with high-cost low-demand customers, with an associated access deficit. There are methods to address the USO problem, but since all regulation will inevitably result in some distortions of incentives, there will be costs associated with these methods too.

In the economic literature, a number of regulatory schemes have been proposed to deal with at least some of these concerns. For a more extensive discussion of these, the reader is referred to the report on which this article is based, but a few aspects of four of the most prominent schemes will be mentioned here.

According to the *Efficient Component Pricing* (or ECP) scheme, the access price should be set so that it is equally profitable for the incumbent to sell access services as to provide the final customers directly. This is achieved – in the simple case – by setting an access price equal to the retail price minus the incumbent's retail costs. The scheme gives the entrant correct incentives to enter the retail market and it gives the incumbent correct incentives to provide access. However, it can create incentives for excessive duplication of the infrastructure and it does not necessarily lower retail prices much below the monopoly level.

To address these shortcomings, Laffont & Tirole (2000) argue in favour of *Global Price Caps*. The idea is to subject the incumbent to an average price cap that applies to wholesale (infrastructural) services as well as to retail services. The task of the regulator will be to set the price cap, but subject to the price cap, all prices will be set by the incumbent. Free to set prices, the incumbent will have incentives to implement a socially optimal price *structure*, although the *level* may be too high. If the price cap as set at the appropriate level, however, the incumbent will have the power to extract enough profits to cover the access deficit and meet possible Universal Service Obligations, while at the same time the price cap will prevent it from setting prices higher than necessary. Hence, in contrast to the ECP scheme, global price caps can provide substantial gains to the consumers. Just as under the ECP scheme, the incumbent has the correct incentives to provide access to rivals, at least in the short run, but a potential problem is that the incumbent is given the possibility (if not the incentive) to eliminate rivals via a price squeeze.

Armstrong (2002) argues strongly that policy makers should make greater use of *output taxes*, levied on incumbents and entrants alike, in order to increase efficiency. The main difference between an output tax and an access fee is that the former has to be paid irrespective of whether the operator uses its own infrastructure or infrastructure owned by someone else, while access fees will only have to be paid in the latter case. The main advantage of introducing output taxes is that they give the regulator one more “instrument” to achieve efficiency, in particular when the regulator (or the legislator) simultaneously pursues other objectives than pure efficiency, such as universal service. Output taxes will then, in fact, be fees to fund universal-service obligations.

Finally, Cave & Vogelsang (2003) have proposed the use of an *escalating access price* scheme. The idea is that access prices should rise over time, in order to stimulate competition in infrastructure. Initially, entrants are likely to rely on access, but competitive investments in infrastructure will become successively more profitable. Although the scheme has its merits and in some respects conforms with the sunset vision, it may seem like a theoretical notion, given that regulated access prices have historically *declined*.

CONCLUSIONS

The choice between competition in services and infrastructure-based competition is a complex one. At the bottom lies the choice between the benefits of free competition and the benefits of returns to scale. Infrastructure-based competition potentially offers less regulation and, hence, less regulation-induced inefficiencies. Service-based competition, on the other hand, allows the industry to realize greater returns to scale. The technological development has reduced returns to scale in infrastructure, but some infrastructural services are still natural monopolies. In addition, regulation is sometimes needed because of returns to scale on the *demand* side (network effects); one example is regulation of call termination. Which choice is the optimal cannot be deduced in the abstract.

One conclusion of this paper is that the vision of a "sunset" for telecom regulation may be at least partially misleading. According to this vision, to which the EU Commission has alluded in the process of launching the E-com directive, sector-specific telecom regulation will eventually become unnecessary: when facilities-based competition has evolved, general competition rules will be sufficient. However, it seems likely that some supply-side returns to scale will persist and even more likely that demand-side effects will continue to require regulation. For example, regulation of call origination is still recommended by the EU Commission for the fixed network (presumably because of supply-side returns to scale) and there is no sign of the regulation of call termination being repealed, presumably because of the large network effects. Consistent with the sunset vision, while the 2003 package recommended ex-ante regulation of access for origination on all networks, also mobile networks, the 2007 recommendation no longer suggests access regulation for origination on networks for mobile telephony. This represents an important shift in policy, from a focus on short-run competition towards a focus on investment incentives.

This brings us to another conclusion, by no ways novel: regulation must be designed so as not to distort investment incentives. In particular, this suggests that while strict access regimes to old monopoly networks may be warranted, one must be careful not to impose too strict access regimes on networks that have been built under competition. In the on-going revisions, a major challenge will be in balancing these two concerns for next-generation fixed networks for broadband services. To what extent should incumbent operators be given strong incentives to upgrade their networks – perhaps by being given "regulatory holidays" – and to what extent should regulation be used to limit the advantage these operators still have from their legacy networks? There is no obvious answer to this question – and no simple answer is provided in the regulatory reform proposal. Instead, the proposal recognises that a good answer requires a careful balancing of the benefits of

good investment incentives and appropriate measures to control the market power that legacy networks create.

More generally, there is no regulatory panacea for natural monopolies. Regulation will always lead to inefficiencies and all regulatory schemes have their weaknesses. In the presence of market power, abstaining from regulation is not a solution either, since the absence of regulation will result in other inefficiencies. There is never a perfect policy, only a "least bad" one, and even that may be elusive. At first glance, facilities-based competition may appear to be the Columbi egg of natural monopolies. A closer look, however, reveals that this method is also imperfect, for the simple reason that duplication is sometimes just too costly, but also because of more complex reasons related to the substantial network effects that exist in the telecom market. But it is important to remember that regulation is the exception and that often the cost of duplication is worth incurring, because it makes competition more intense and because it reduces the regulatory burden.

REFERENCES

Armstrong M. (2002), "The Theory of Access Pricing and Interconnection" in M.E. Cave M.E, Majumdar S.K & Vogelsang I. (eds.), Handbook of Telecommunications Economics, Vol. 1, North-Holland, Amsterdam.

Bergman M. (2002), "Lärobok för Regelnissar", En ESO-rapport om regelhantering vid avreglering, Ds 2002:21, Stockholm.

Bergman M. (2008), "Competition in Services or Infrastructure-based Competition?", manuscript available at www.sh.se/matsbergman (an earlier version was published in An anthology on the foundations for competition and development in electronic communications markets, PTS, Stockholm, 2005.)

Cave M.E, Fogelsang I. (2003), "How Access Pricing and Entry Interact", Telecommunications Policy, Vol. 27, pp 717-727.

Laffont J.J., Tirole J. (2000), "Competition in Telecommunications", MIT Press.

PTS (2003), "Svensk telemarknad första halvåret", report PTS-ER-2003, December 18th

The Arbitrage between Service and Infrastructure Competition in the Light of the "Ladder of Investment" Theory[12]

ANDERS HENTEN
KNUD ERIK SKOUBY
Center for Communication, Media and Information technologies (CMI), Aalborg University

***Abstract**: During the past few years, a debate has erupted on the theory of 'the ladder of investment' in the broadband access area. The ladder of investment theory says that the entry of players onto the market takes place step-wise ('rung-wise') and that players will enter the market in the areas where barriers to entry are low and then, later in the process, will enter areas with higher barriers. In the telecommunications field, this means that new operators will first enter as service providers, who do not need to make as substantial up-front investments as infrastructure providers, and then later on move to infrastructure investments. The question is whether this is right or whether the theory put into practice in regulatory policy is counterproductive to the development of infrastructure and infrastructure competition. This is the topic of the present paper discussing the theory and its applicability and consequences. First, there is a frame-setting discussion on the theory and the environment in which it has developed. After this, three country cases (the UK, the US, and Denmark) are briefly presented with an emphasis on how these countries have approached the issue of facility based vs. service based competition. Third, there is an analysis and conclusion putting together the outcome of these case stories and the analytical framework.*

INTRODUCTION

In Europe, during the past few years, a debate has erupted on the theory of 'the ladder of investment' in the broadband access area. Not that the theory is confined to the broadband access area; but being an area subject to much focus lately and with a continued dominant position of the incumbent fixed line operators, the discussion has centered on broadband access.

The ladder of investment theory says that the entry of players onto the market takes place step-wise ('rung-wise') and that players will enter the market in the areas where barriers to entry are low and then, later in the process – when they have established themselves on the market – will enter areas with higher barriers. In the telecommunications field, this means that new operators will first enter as service providers, who do not need to make

12 *This paper is an edited and shortened version of a paper presented at the 16th Biennial ITS conference in Beijing, China, on 12-16 June 2006.*

as substantial up-front investments as infrastructure providers, and then later on move to infrastructure investments.

This constitutes the general aspect of the theory and applies to markets in general: it is based on the concept of barriers to entry and can empirically be witnessed in a great deal of markets. The more special aspect regarding telecommunications is that telecommunications has come from a situation with legal monopolies: that the incumbents still dominate after the liberalization of markets and that there is a political sentiment that this is counterproductive to the development of telecommunications and, therefore, not acceptable – even though quasi-monopolies also are known from other markets. Especially the access networks constitute a problem, as they are difficult to replicate in a competitive manner. Therefore, regulation is necessary – is the argument. Regulation thus becomes part of the ladder concept in the sense that regulation must help alternative operators to enter the market so that they can start climbing 'the ladder of investment'.

There are basically two camps in the discussion. In the one camp, one finds the incumbent fixed line operators and their European organization, ETNO (European Telecommunications Network Operators' Association). They believe that the theory of the ladder of investment is harmful, that there should be less network access regulation (and regulation in general), and that facility- (infrastructure)[13] based competition should be given priority over service-based competition. In the other camp, one finds the operators that use the networks of the incumbents to get access to end users and their European organisation, ECTA (European Competitive Telecommunications Association). They believe that the ladder of investment theory is useful, that network access regulation is necessary to secure competition, and that service-based competition is one of the ways to stimulate alternative access.

The fact that the arguments put forward are somnambulistically predictable does, however, not means that they are unreasonable and not worth discussing. But it is important to note the vested interests in the different positions, and the question is whether there is a 'right' position from a societal point of view.

Can it be determined whether the ladder of investment theory is helpful from a societal viewpoint? This is the topic of the present paper discussing the theory and its applicability and consequences. First, there is a frame-setting discussion on the theory and the environment in which it has developed. After this, three country cases (the UK, the US, and Denmark) are briefly presented with an emphasis on how these countries have approached the issue of facility based vs. service-based competition. Third, there is an

13 *In the paper, the terms facility and infrastructure are used interchangeably. The same applies to service based and access based competition.*

analysis and conclusion putting together the outcome of these case stories and the analytical framework.

THEORY FRAMEWORK

The texts most often referred to when discussing the theory of the ladder of investment are written by Martin Cave (and co-authors)[14]. The basic idea is that entry of alternative operators is easier in some parts of the value networks of supply than in others, i.e. there is a difference in the degree of replicability – or barriers to entry. As stated by Cave: 'replicability is not a simple binary variable'[15], and a differentiation must be made between more or less replicable assets. It is, for instance, easier to replicate access to an IP network than replicating a local loop. And, to illustrate the matter, Cave presents a ladder of replicability of assets, going from retailing, IP network, backhaul, DSLAM and ending with local loop[16].

The concept of the ladder of investment is strongly supported by the European organisation for competitive telecommunications operators, ECTA. In a document from 2005, 'ECTA concludes that if action is taken by regulators to promote competition by providing certainty about the ladder of investment, improving the product offerings available to competitors and combating discrimination by the incumbents in favour of their own broadband retail offers, there should be further scope to increase choice for consumers and boost broadband penetration'[17]. In their mind, there is no doubt that the implementation of the ladder of investment in regulatory provisions is beneficial to society – not to mention their own interests. ECTA emphasises the promotion of competition, the choice for consumers between operators, and broadband penetration.

ETNO, the European organisation for incumbent fixed line network operators, on the other hand, is convinced that 'the ladder concept does not create the prospect for increased investment in next generation access infrastructures and inter-platform competition in Europe'[18]. Again, there is no doubt that the critique of the ladder concept serves the interests of the fixed line incumbent network operators. But they also have an argument viewing the matter from a societal point of view, namely that the ladder concept, in their opinion, is counterproductive for investments and for the establishment

14 *For instance, Cave & Prosperetti (2001), Cave & Vogelsang (2003), Cave (2004).*

15 *Cave (2004) p. 8.*

16 *Ibid. p. 20.*

17 *ECTA (2005). p. 2.*

18 *ETNO (2005), p.1.*

of inter-modal competition. Who wants to make costly and risky investments if the new access infrastructures are accessible for all other operators, is the argument.

From the regulatory side, the theory of the ladder of investment concept largely finds support. The European Regulators Group for electronic communication networks and services (ERG) has in different documents expressed a positive attitude towards using the concept[19]. In the same vein of cynicism as with respect to the ECTA and ETNO positions, one could see this as an expression of the potentially many regulatory activities involved in using the ladder of investment concept. However, it is much more productive to see it as part of the larger regulatory legacy and concept of telecommunications regulation in Europe. The Open Network Provision (ONP) policy, which was developed, at first, when liberalising the telecommunications markets in Europe, was largely based on a competitive model with service competition as the primary form of competition – which was quite understandable taking into consideration that the European telecommunications markets (as elsewhere) came from a monopoly situation and that the incumbents were in possession of almost all access paths.

As compared to the former ONP based regulation, the present EU regulatory framework implemented in 2003 is more focused on infrastructure competition, and the tendency is, furthermore, to limit sector specific regulation where it is considered possible. The trend is thus towards infrastructure-based competition, and the position of the incumbent fixed line network operators can best be understood as a further pressure in this direction and a concern that the process is not moving fast enough. It would be difficult to defend the position that regulation in the EU with the existing regulatory framework is heading in direction of more service competition and that the ladder concept is an expression of such a direction. The whole idea in the ladder concept is to move competition towards infrastructure-based competition – maybe not fast enough, but the direction is clear.

With the latest propositions (2007) from the European Commission[20] to change the regulatory framework, the concept of functional separation is introduced. As opposed to structural separation, functional separation does not mean an ownership separation of incumbent operators. It means that the infrastructure part and the service part are separated accounting-wise and management-wise. Such a separation can be seen to promote service competition, as the purpose would be to help create a level playing field for all operators offering services on the basis of the infrastructure of the

19 *ERG (2004), ERG (2005).*

20 *See European Commission, "Legislative proposals" http://ec.europa.eu/information_society/policy/ecomm/library/proposals/index:en.htm*

incumbent including the incumbent itself. From the part of the European Commission, however, functional separation is only foreseen to be a last-resort regulatory tool – if nothing else works.

In a paper from 2004[21], Alison Oldale and A. Jorge Padilla strongly oppose the investment ladder theory and their major claim is that the proponents of the ladder theory do not see any contradiction between service and infrastructure competition: the ladder reconciles the trade-off between service and infrastructure competition – as they state it[22]. Their point of view is that if service competition is made attractive for alternative operators, they will not invest in their own infrastructure, and the incentive for existing operators to upgrade their networks will be lowered. They, therefore, believe that service-based competition leads to a kind of static efficiency while infrastructure competition leads to dynamic efficiency.

Their major contention, though, is related to the regulatory presumptions that they see behind the ladder of investment theory. As the regulatory idea in the ladder theory is that regulators should create incentives for operators to climb up the ladder, Oldale & Padilla state that this must hinge on, first, a presumption of regulatory omnipotence and, secondly, a presumption regarding the sustainability of a fragmented market. With respect to regulatory omnipotence, their point is that intervention in the market, if it is to have positive effects, requires an optimistic view of the resources, information and competence available to regulators[23]. Regarding sustainable fragmentation, the issue is that they believe access based competition will lead to a fragmented market and that this will require perpetual regulation[24].

The question is whether these alleged presumptions really characterise the ladder theory. One will hardly find any proponent of the ladder concept who will subscribe to the opinion that regulators are omnipotent. The point of view is that there are problems in the markets (e.g. dominance) that need to be rectified by means of regulatory intervention in spite of regulatory failures partly caused by lack of resources, information and competence. The point regarding perpetual regulation as a result of fragmented markets is only valid if it is accepted that there will not be a move towards forms of competition involving an increasing degree of infrastructure assets and that the implementation of the ladder concept fixates the market in a service-based competition mode. This is one of the main controversies in the debates on the ladder of investment. On the one hand – Oldale & Padilla, for instance – take the position that there is a clear trade-off between access based entry

21 *Oldale & Padilla (2004).*

22 *Ibid. p. 69.*

23 *Ibid. p. 71.*

24 *Ibid. p. 73.*

and a sustainable facility based competition. Cave, on the other hand, makes the point that it is not a binary choice, but that there are intermediary stages between pure service-based competition and pure infrastructure-based competition. This controversy, however, cannot be finally settled at a purely theoretical level. The following section will, therefore, look into the actual developments in three national markets, the UK, the US, and Denmark.

COUNTRY CASES

➢ *United Kingdom*

The UK was among the first of countries to embark on a policy of liberalisation in the telecommunications area. There was a duopoly period from 1984 to 1991, which was followed by a period up to 1997, where the market was opened to further competition. Both periods can be seen as having an emphasis on infrastructure competition. Since 1997, there has been more focus on service competition and a more balanced approach to encouraging infrastructure as well as service competition[25]. The UK has thus been seen as the country in Europe with most focus on infrastructure competition but has moved towards the European 'mid-field' trying to promote different kinds of competition on the market at the same time.

In the 'Strategic review of telecommunications – phase 2 consultation document' from 2004, OFCOM suggested 7 regulatory principles[26]:

- Promote competition at the deepest levels of infrastructure where it will be effective and sustainable.
- Focus regulation to deliver equality of access beyond those levels
- As soon as competitive conditions allow, withdraw from regulation at other levels.
- Promote a favourable climate for efficient and timely investment and stimulate innovation.
- Accommodate varying regulatory solutions for different products
- Create scope for market entry which could, over time, remove economic bottlenecks.

25 *Whalley (2005), p. 136.*

26 *Whalley (2005), p. 138.*

- Adopt light-touch economic regulation based on competition law and the promotion of interoperability, unless there are enduring economic bottlenecks.

These seven principles very well exemplify the balanced 'mid-field' approach that OFCOM has chosen to take regarding promoting broadband competition: Encouraging infrastructure competition where possible and creating a level playing field with equality of access.

➢ ***USA***

In 1996, the Telecommunications Act was introduced in the US with the clear aim to enhance and introduce new competition in telecommunications provision. Three forms of competition were envisioned: facility-based competition, service-based competition, and unbundling as a hybrid, where operators purchase components from the incumbents and combine them with own their facilities. The latter two forms were seen as transitory and the first as representing the real competitive situation. From this, it is quite clear that the intention of the 1996 Act was to stimulate facility based competition as the long term solution and enable the two other forms through regulatory measures on a market with high barriers to entry.

From 1999, as competitive conditions in the US became more difficult as a result of the telecommunications collapse, the FCC expanded the unbundling to include leasing by competitors of the incumbents' entire local service - known as UNE-P (Unbundled Network Elements Platform) - based on TELRIC (Total Element Long Run Incremental Cost) prices. The result was that infrastructure-based competition stagnated while the proportion of lines based on UNE-P came to account for nearly half the total[27].

UNE-P thus became the preferred market entry method, but there seems to be general agreement that the system influenced unfavourably the development of broadband and advanced networks in the US compared to, e.g., countries in Europe and Korea and Japan that have taken the lead in this development[28]. The argument is that TELRIC has disadvantaged the incumbents and, therefore, weakened them as potential investors in advanced infrastructures.

In 2005, an Order was released by the FCC that redefined the unbundling obligations of the incumbents and drastically reduced these obligations

27 *Cave (2004), p. 10.*

28 *Bauer (2005), p. 161.*

especially in relation to advanced networks. Also, the incumbents were to a great extent allowed to charge market prices for the network elements leading to prices typically 300-500 % above the TELRIC prices[29].

- ***Denmark***

While Denmark in the first phases of the telecommunications liberalisation processes in Europe was not to be found among the first-movers, the Danish government and parliament made a turn in 1994/1995 and decided to liberalise the Danish telecommunications sector one and a half year ahead of the EU 1998 deadline, i.e. in the summer of 1996. The slogan and title of the policy document constituting the basis for the Danish liberalisation process was 'Best and cheapest by way of real competition'[30]. Focus was on promoting competition as fast as possible, and service competition was seen as the most feasible way to get started.

In 1999, a policy agreement was made in parliament with the most important goal being to 'promote access to the network society' and the most important means being to 'increase competition in the access market'[31]. Simultaneously, the Danish NRA started promoting the idea of 'several pipes to the home', pointing not only to the variety of different technologies, which can be used for accessing telecommunications services, but also to the importance of access infrastructure competition.

It would be a misinterpretation to conclude that priority has been given to infrastructure access competition since 1999. It is more correct to say that both kinds of access competition have been promoted and that the policy has been to 'walk on two legs' - although this expression has never been used.

In a status on the development of broadband access in Denmark it is emphasized that 'the national broadband strategy is based on a market-driven infrastructure development, *facilitated by an opening of access to competitors through interconnection agreements and by encouraging rollout of several, open and competing broadband access "pipes to the home*"' (our italics)[32].

29 *Bauer (2005), p. 157.*

30 *Ministry of Research (1995), "Best and cheapest by way of real competition".*

31 *Telepolitisk aftale (1999).*

32 *National IT and Telecom Agency (2004) "Mapping of Broadband Access Services in Denmark –Status by mid-2004", December, p. 2.*

An overall strategy in the field of infrastructure vs. service competition has never been explicitly formulated by the political and regulatory authorities in Denmark. Statements concerning the necessity of supporting competition in the infrastructure as well as the service fields are common and the importance of infrastructure competition as a more sustainable form of competition in the long run, decreasing the need for sector specific regulation, is recognised. But there is no overall priority given to infrastructure competition, and positions have not been taken on the issue, e.g., as to whether the promotion of service competition is a barrier to infrastructure competition or can be seen as a stepping stone towards infrastructure competition. The Danish policy and regulatory approach in the field can be characterised as pragmatic, promoting infrastructure competition when this is seen as possible and advantageous but also seeking to support service competition, making it possible for competitive suppliers to gain market shares.

ANALYSIS

Unfortunately, the empirical evidence presented in this paper – the three case stories – is not more conclusive than the theoretical discussion[33]. The development in the UK is a case story of a country that moved from a strict focus on infrastructure competition towards putting more emphasis on service competition. The US, on the other hand, has moved in the opposite direction. Whilst the Telecommunications Act from 1996, indeed, had a primary emphasis on infrastructure competition, the focus changed around the turn of the century as a result of the crisis in the telecommunications industry and the FCC decision to include the leasing of the incumbents' entire local service into the unbundling requirements. This tipped the marked in favour of service-based competition, and the FCC has lately changed the unbundling policy in order to promote infrastructure-based competition. Denmark, finally, is an example of a country which has tried to balance the promotion of service and infrastructure-based competition – as the UK also has done in recent years.

One obvious conclusion is that there are many different ways to expand broadband take-up. It depends on the specific national contexts including the stage and kind of development hitherto experienced. In the UK, the policy had to be adjusted, as the strong focus on infrastructure competition had led to a slower broadband diffusion than in comparable countries. In the US, the policy had to be adjusted in the other direction taking into consideration that

33 *The same applies for the material presented in ERG (2005). This document aims at documenting the advantages of implementing a ladder of investment approach, but the evidence presented is not conclusive.*

the UNE-P policy had led to a halt in infrastructure competition. There is, therefore, no one correct broadband access policy and countries that canonise one kind of policy will be likely to encounter serious problems if policies are not adjusted to the specific circumstances.

For a period of time, it was mostly the ladder concept being canonised. It has a strong appeal, as it is easily understandable and seems to promise a simple formula for success. This, in itself, must lead to scepticism. However, there is no doubt that a 'ladder-driven' policy starting from the lowest rungs of the ladder with an emphasis on service competition, in most cases, gives a faster growth in competitors' market shares. In all probability, it therefore also provides a faster take-up of broadband access, as there will be a pressure on reaching customers and on lowering prices.

What is the problem in this? One problem can be lower investments in new infrastructure - as stated by the opponents of the ladder theory. There will be operators considering investing in alternative infrastructures who will hold back their investments if they cannot recoup their investments sufficiently fast when prices are low because of service competition. On the other hand, if there is no immediate likelihood that alternative infrastructure investments will be made, it is better to kick-start the market by introducing attractive offers for alternative service providers.

This is the balance that has to be struck, and if one has no trust in the potential positive outcomes of regulatory intervention (as in the case of Oldale and Padilla 2004), the obvious conclusion is not to promote service-based competition, as it cannot, in this understanding, be determined when it is right to adjust the policy. Even if it is admitted that regulators may have understood correctly when to intervene in order to move the market to higher rungs of the ladder of investment, it creates an uncertainty for service providers if they cannot count on the continued existence of the market conditions on which they made their market entry – which is why ECTA emphasises 'providing *certainty* about the ladder of investment' (out italics).

It, therefore, comes down to trust or lack of trust in the positive outcomes of regulatory interventions. If there is no trust in regulatory intervention in general, there can be no trust either in regulatory interventions based on a ladder of investment concept. But there are two other aspects that have to be considered when determining the usefulness of the ladder concept. The first thing is that even though a simple interpretation of the ladder concept assumes that operators, as time goes by and as they are led in an upwards direction, will climb up the ladder, there will in fact often be operators staring from all 'rungs of the ladder'. Some companies will, at any rate, focus on infrastructure competition and will not consider being service providers. This applies, for instance, to electricity utilities. Other companies will only very reluctantly move up the ladder, as their whole market strategy is service-based. A too simplistic view on the market, which does not consider the

fundamental differences in strategies of companies, will fail to understand the possibilities in a multi-tiered (multi-rung) regulatory policy. It is, as emphasised by Cave, not a binary choice.

Secondly, and from a different perspective than merely a competition based discussion, the infrastructures of the incumbents can also be considered as societal infrastructures that should be at some kind of disposal of different service providers in society. From this perspective, the discussion on access to network infrastructures resembles a number of other similar discussions on the advantages of private property vs. the benefits of shared access to societal resources. A case in point deals with software patents. Should software patents and the kind of private property rules that they entail be promoted or should software patents be disallowed and a greater degree of shared use of software be supported? Just as inconclusive as this discussion, is the discussion in the present paper regarding the societal benefits of the promotion of service competition and the ladder of investment concept. The interest of the different market players is clear. However, the societal level is less obvious.

CONCLUSION

The paper examines the issue of service and/or infrastructure competition in light of the 'ladder of investment' concept. Emphasis is on the arguments in favour of service and/or infrastructure competition when discussing the development of access competition. Furthermore, country cases are included as illustrations of differences in development paths.

The opponents of the ladder concept have taken a very fundamentalist view on service and infrastructure competition: Service competition is seen as a substitute for infrastructure competition and not as a complement. The proponents of the theory, on the other hand, recognise that infrastructure competition is more sustainable than service competition, but believe that ways into building up competition must be established in a situation where it is difficult for alternative operators to enter the market in direct infrastructure competition with the incumbents.

The incumbent operators have the strongest case with respect to new infrastructures. Even if incumbents get their costs covered (using appropriate costing methods), there will still be an investment risk, which is not taken into consideration and which may hold back investments if the potential competitive advantages cannot be retained. The case is much weaker regarding legacy networks, which can be considered as societal infrastructures which should be open to all service operators. When differentiating the regulation of new and old infrastructures, there will also be an incentive for incumbent operators to invest in new infrastructures. The

intentions of promoting competition and penetration as well as investment and innovation can thus, to some extent, be reconciled.

The other argument against the ladder concept is that it presupposes a level of information and competencies at the regulatory institutions that does not exist. The twists and turns of the regulatory policies, having been witnessed in the UK and the US, could be taken as proof of the impossibility of consistent regulatory interventions. However, they can also be seen as evidence of the ability to adjust to the ever changing developments. Not that the UNE-P policy in the US and the former one-sided focus on infrastructure competition in the UK should be seen as appropriate policies in their specific situations. They are most likely cases of regulatory failures. Furthermore, they are a warning against canonising a specific policy and establishing it as an eternal 'truth'. It is, for instance, difficult to see how an entirely infrastructure-based competition policy would have led to a situation in the broadband access markets with more competition, higher penetration, more investments, and more innovation. And, it is also difficult to see that this will be the case in the immediately coming years.

REFERENCES

Bergman M. (2004), "Competition in Services or Infrastructure-based Competition?", Swedish National Post and Telecom Authority, Stockholm.

Bourreau M., Dogan P. (2004), "Service-based vs. Facility-based Competition in Local Access Networks", Information Economics and Policy, 16, pp. 287-306.

Cave M. (2004), "Making the ladder of investment operational", Paper presented to the European Commission, November.

Cave M., Prosperetti, L. (2001), "The European Telecommunications Infrastructures", Oxford Review of Economic Policy, vol. 17, issue 3, pp. 416-431.

Cave M., Vogelsang I. (2003), "How Access Pricing and Entry Interact", Telecommunications Policy, 27, pp. 717-727.

Copenhagen Economics (2005), "Prisen på rå kobber: Principielle overvejelser om fastsættelse og fordeling af prisen på rå kobber, IT- og Telestyrelsen", Copenhagen, Denmark.

DIW Berlin (2005), "Infrastruktur- vs. Dienstewettbewerb: Regulierungspraxis bei Teilnehmeranschlussleitungen im europäischen Vergleich", DIW, Berlin, Germany

ECTA (2005), "Leading Broadband Countries Benefit from Competition".

ERG (2004), "Common Position on the Approach to Appropriate Remedies in the New Regulatory Framework", ERG (03).

ERG (2005), "Broadband Market Competition Report", ERG (05) 23.

ETNO (2005), "Reflection Document on Re-Assessing the "Ladder of Investment" in the Context of Broadband Access Regulation".

Henten A., Skoubi K.E. (2005), "Regulation of Local Loop Access – Infrastructure versus Service Competition in Fixed Broadband Access: The Case of Denmark", ITS Europe Conference, September 4-6, Porto, Portugal.

Oldale A., Padilla A.J. (2004), "From State Monopoly to the "Investment Ladder": Competition Policy and the NRF", in Konkurrensverket: "The Pros and Cons of Antitrust in Deregulated Markets, Konkurrensverket", Stockholm, Sweden.

Infrastructure-Based versus Service-Based Competition in Telecommunications: An Economic and Regulatory Review of National Implementation Strategies

JÖRG KITTL
Senior consultant at JUCONOMY Consulting AG

MARTIN LUNDBORG
Senior consultant at JUCONOMY Consulting AG

ERNST-OLAV RUHLE
CEO at JUCONOMY Consulting AG

***Abstract**: Unbundling of the local loop (ULL) has seen different "success stories" across Europe. Although the obligation for the provision of ULL was early implemented in the regulatory framework national, implementation has been heterogeneous. National Regulatory Authorities (NRAs) wondered whether to foster service-based or infrastructure-based competition. This paper analyses whether the strategies of NRAs have had any effect on the economic welfare. It indicates that infrastructure-based competition is important for business customers and has a positive effect on innovation. Service-based competition lowers call prices and is more important for residential markets. The study points out the importance of a balanced approach to both types of policies.*

INTRODUCTION AND TYPES OF COMPETITION

The issue researched in this paper goes back to the questions discussed when markets for telecommunications services were opened up, namely what kind of competition delivers more favourable results: infrastructure or service-based competition?

Regulatory policies in numerous countries have developed in very different ways. The US started with unbundling local loops (ULL)[34] rather early and introduced also service-based competition[35]. European countries followed an almost joint approach to market opening with the 1998 framework[36]. Many

34 *Crandall (2005), p. 7.*

35 *Vogelsang (2002, 2005).*

36 *See the implementation reports from the EU Commission: http://europa.eu.int/information_society/topics/telecoms/implementation/index_en.htm*

incumbents criticized market opening with service-based competition as arbitraging[37]. Access competition through ULL was compulsory only after the year 2000[38]. With that step, further business models have arisen[39]. The move towards all-IP/NGN networks extends the debate today to net neutrality[40].

According to the European framework efficient investment in infrastructure shall be encouraged and innovation promoted[41]. This regulatory framework already offers several ways of fostering competition (toolbox) and regulators have chosen different strategies. The main point in discussions has been whether infrastructure or service-based competition would lead to lower prices, more differentiated and innovative products and improved services for consumers. The cost of regulation is an additional point[42].

A frequently discussed concept is the ladder of infrastructure competition, which argues that new entrants may enter the market based on a wholesale product where they only cover minor elements of the value chain (such as resale) and then move on to "higher rungs" of the ladder. Hence, by implementing this ladder, both infrastructure and service-based competition are promoted[43].

37 *Piepenbrock & Schuster 2003 (eds.), Anreize für Infrastrukturinvestitionen bei der Zusammenschaltung in der Telekommunikation.*

38 *Regulation No. 2887/2000 of the European Parliament and of the Council on Unbundled Access to the Local Loop: http://europa.eu.int/ISPO/infosoc/telecompolicy/en/regullfin-en.pdf.*

39 *Cave (2004). It seems that some countries are following the approach of the ladder of investment. See ERG (05)23, 2005, Annex A "Country Studies": http://erg.eu.int/doc/publications/erg_05_23_broadbd_mrkt_comp_annex_a_p.pdf*

40 *See Nicholls (2006), Von Schewick (2007).*

41 *See, Art.8 §.2 of the Framework Directive: Directive 2002/21/EC of the European Parliament and of the Council of March 7th 2002 on a common regulatory framework for electronic communications networks and services (Framework Directive), Official Gazette, no. L 108/33 of April 24th 2002.*

42 *Ellig (2005).*

43 *Vogelsang (2005), p. 58.*

EMPIRICAL ANALYSIS

In order to answer the question of the study, we first examined which countries followed which type of policy. This paper focuses on the EU-15 countries (belonging to EU before May 1st, 2004) since they all had the same regulatory framework for a longer period of time. The categorization between service and infrastructure-based competition in this paper is based on the principle that the replication of the local loop does not make sense[44] and therefore the last mile is a relevant differentiating factor between service and infrastructure-based competition[45].

We divided the countries into two different clusters based on statistics regarding the success of the different regulated products and the date of introduction for these products.

- Group 1: infrastructure-based competition has been the main objective of the NRA visible in its policy and decisions.
- Group 2: the NRA has focused on service-based competition to a greater extent.

➢ *Market development indicators*

For clustering, we look at main indicators describing the performance of competition in order to identify the type of policy NRAs have pursued. For the empirical study, it is assumed that NRA have followed intentions regarding service-based and infrastructure-based competition and that these intentions have had an impact on market development.

The first indicator for the clustering assesses how successful infrastructure-based competition has been by describing the competitors' markets shares with respect to access line competition and the proportion of service-based competition (resale and bit-stream access) of all wholesale products.

---

44 Cave (2004) p.8. This assumption may be debated in light of the NGN discussion (see Von Schewick 2007) and the development of competition in the USA in recent years after major parts of the unbundling regime was abolished (see Crandall 2005).

45 This demarcation is often contested. The Danish regulator, considers only the full replication of the whole infrastructure including the last mile as infrastructure-based competition. See: Henten & Skouby (2005), p. 2.

The second indicator uses the competitors' market shares of local loops (both, line sharing and fully unbundled loops) and the market shares of all competitors in broadband markets (measuring the success of competitors).

The third indicator measures the success of competitors in broadband markets in relation to their share of the retail market. By looking at the proportion of infrastructure-based competition and thereby excluding CATV, the same analysis can be conducted for the DSL-market only, which gives us the fourth indicator.

- ***Regulatory policy indicators***

Secondly, we examined the date at which different wholesale products were introduced. We used the dates of introduction of ULL and carrier (pre)selection services as the fifth indicator.

- ***Results of the quantitative clustering***

Summing up the different indicators described above gives the following clustering of the countries (see also Annex I for detailed figures)[46].

Table 1. Topology of national regimes

Infrastructure-based policy	Service-based policy
Germany	Austria
Denmark	Spain
France	Italy
Netherlands	Ireland
Sweden	UK
Finland	

46 Four countries (Portugal, Belgium, Luxembourg and Greece) had to be excluded as the outcome was not clear.

THE EFFECTS OF REGULATORY POLICY ON THE COMPETITION SITUATION

➢ *Price and penetration as measurements of competition and innovation*

The main aim of liberalization in the EU was to increase overall welfare through lower prices, enhanced consumer choice, innovative products, etc.[47]. Therefore, we examined the welfare effects measured by the state of competition, which is defined through the price situation. This is based on the assumption that more competition reduces prices in the market.

Competition can also increase consumer welfare without reducing prices. This is achieved by innovation. We assume that innovation and penetration rates of new services and technologies correlate[48]. Therefore, we measured innovation by the penetration rates of broadband uptake, as well as the uptake of ISDN[49].

The statistical method chosen is the heteroscedastic t-test[50], as the variance of the two clusters have different values for the variances.

➢ *Effects on competition and innovation*

Based on the indicators for penetration rates, price competition, price development from 2000-2004, the outcomes on competition and innovation were calculated. The results are shown in the tables in Annex.

We conclude that prices are lower in those countries with predominantly infrastructure-based competition[51]. Furthermore, there are higher penetration rates on average in countries with infrastructure-based regulatory policies.

---

47 Commission Staff Working Document, "Europe's Liberalised Telecommunications Market - A Guide to the Rules of the Game", from October 18th 2000. http://europa.eu.int/ISPO/info-soc/telecompolicy/en/userguide-en.pdf

48 See, for example, McNary (2001), ERG (2005b).

49 See, for example, Von Schewick with respect to different forms of innovation and their assessment with regard to the discussion on network neutrality.

50 Confer to http://www.statsoft.com/textbook/stathome.html.

51 Remark: the variances are relatively large due to the small sample. However, by comparing the means and the variances, the results are strong and, in several cases, achieve a confidence interval of 95%. So, the results are to be seen as significant.

These results are only significant for broadband markets, but not for ISDN[52]. In addition, the pricing effects are more significant for the business customers in countries with infrastructure-based competition.

With regard to price levels in 2004, these are lower in countries with infrastructure-based policies. The results, however, are not significant for the residential baskets. The differences between the clusters are especially high for business customers.

Result 1: *Countries with predominantly infrastructure-based competition have lower overall prices and thereby more innovation.*

Result 2: *Analysis of penetration rates tends to indicate that infrastructure-based policies foster higher penetration rates.*

Result 3: *There is an indication that infrastructure-based competition is of greater importance to business customers than to residential customers.*

When looking at the prices in 2005, however, the results are no longer significant. Prices do level off, especially for residential customers, but also for business customers. The changes that took place between 2004 and 2005 are elements of a long-term process. Hence, a trend towards price harmonization did emerge between the two clusters during the period 2000 to 2005 – although not all results are significant.

Result 4*: Price differences between countries are diminishing in Europe – irrespective of the main competition form.*

In the period 2000 to 2004, the monthly rental went up by almost 30% on average in countries with service-based competition, while prices in countries with regulation focusing on infrastructure-based competition increased by only 3.9% (business customer) and 8.8% (residential customers). The increase in monthly rental prices was over-compensated by far in those countries with service-based competition. In fact, the prices for OECD baskets (which also include monthly rental) decreased by 12-16% over the same period in countries with service-based competition. In those countries with more infrastructure-based competition, OECD baskets

52 *There is a risk of auto-correlation in this case, since the countries were clustered partially by the penetration rates, but as other factors were considered as well, this risk has been reduced.*

decreased only slightly. This indicates that service-based competition is more important for residential clients, while infrastructure-based competition is more important for business customers.

Result 5:	*There are clear indications that tariff rebalancing has gone further in the countries with service-based competition than in those with infrastructure-based competition.*

- ***Critical remarks***

A critical remark has to be made regarding the problem derived from the role of CATV, which has existed all of the time and has had a positive impact on competition, but has hardly been influenced or promoted by regulation. Therefore, assessing the impact of this alternative infrastructure with intermodal competition correctly is difficult.

Moreover, the sample sizes are rather small and the variances in several cases are large. The results in this study consequently need to be considered with care and it is advisable to verify the results in future research.

ANALYSIS AND DISCUSSION

- ***Types of competition***

The empirical study has shown that infrastructure-based competition has led to significantly lower access costs and call tariffs for business customers and, to a slightly lesser extent, for residential customers in its early stages. However, call tariffs, as well as access costs for fixed line services (monthly line rental) have risen after a decrease for both customer groups slightly again over time in countries with infrastructure-based competition.

On the other hand, where service-based competition is fostered, infrastructure-based operators had compensated the loss in turnover from call prices by sharply increasing line rental prices (either by tariff rebalancing and/or by including call prices in the price of the rental - bundling). Sharply reduced call tariffs mostly overcompensated for the line rental increases. Compared to infrastructure-based countries, call prices saw an overall steeper decline in service-based countries.

In recent periods, differences in the overall price structures in service-based and infrastructure-based countries tend to diminish. This can be traced back

to the fact that NRAs started to implement the ladder of investment step by step. The benefits of both strategies are now beginning to evolve and compensate for the negative aspects of the other – reduced access costs are linked to low call tariffs.

Conclusion:	*With a stepwise introduction of the ladder of investment NRAs can focus on a specific liberalization strategy. But only complete implementation of the ladder of investment will bring positive welfare effects.*

Competitors will undertake investment in their own infrastructure only if the return justifies that investment. Therefore, business customers with a high impact on return on investment are those entities that benefited most from infrastructure-based competition. Due to sunk costs and economies of scale, only a limited number of ULL-operators is able to survive in a market place (oligopoly). In such a situation, ULL operators have strong incentives to collude and raise prices, as the locking effects for consumers are significant[53].

In this situation, service-based competition could help to keep prices low if introduced as a complement.

Conclusion:	*The complementary introduction of service and infrastructure-based competition limits the negative outcomes of either and supports the development of the positive elements of both liberalization strategies.*

Although, intramodal competition is not reflected in this study, one has to bear in mind the effects of such competition. Especially the different cost structures and the different technologies involved create some room in which competition can develop. NRAs need to be aware of the regulatory interdependence created by intermodal and intramodal competition. If one side of competition is regulated, but the other side is not, the question arises whether this leads to a distortion of competition (such as the non-regulation of CATV-networks or non regulated mobile access networks)[54].

---

53 On the issue of complementary products and foreclosure in the NGN world see Von Schewick (2007).

54 A similar question was raised in a decision by the Dutch Competition Court, which annulled an NRA decision on mobile termination market analysis and remedies.

***Conclusion**: Technology neutral regulation is a means of limiting the risks of distortions of intermodal and intramodal competition.*

- ***Management decisions and their influence on investment***

As far as the influence of the NRA's decisions on how to open up the market to competition is concerned, the influence of operators must not be overlooked, as it is them who decide on the investments. The following list contains key factors that influence management decisions.

The liberalization period was rather short, making time to market a key factor. In countries like the UK, France or Germany it was evident that once a clear decision was taken by the NRA on a specific access product, market players focused on that product immediately[55]. From this, we can derive the importance of the first decisions made by NRAs regarding access products for the market.

During the liberalization process, the first goal of a company was to gain market share. The easiest way of achieving this was to duplicate the incumbent's products and offer similar products at a reduced price. Therefore, it is sometimes proposed to set lower prices for products at lower rungs of the ladder of investment. NRAs then raise prices for low rung products of the ladder over time to incentivise investments in access products of higher rungs and to "force" competitors to invest in infrastructure. Yet one has to bear in mind that this proposal may distort competition for late entrants. Climbing the ladder of investment should be a possibility at all times as competition may not necessarily be carried out via pricing, but also via product differentiation.

In telecommunications, mainly new technologies enable innovative services. Infrastructure based operators form the basis of such innovations as they have control over the development and use of the infrastructure. Wholesale obligations may even foster competition, penetration rates and the introduction of innovative services and technologies. Investments in innovative services will be made when the access obligations and thus the returns on investment of the wholesale business are neutral (economically) compared to the retail business.

55 *Bergman (2004).*

Conclusion:	*Market liberalization follows the clear path of a product life cycle in competitive markets. Regulation paves the way for investors to enter markets and the products they intend to introduce.*

- ***Pricing***

To be able to make the ladder of investment operational it is necessary that prices in wholesale markets for the different products are consistent but also that there are clear rules for migration from one product to another[56].

> "*The imposition by national regulatory authorities of mandated access that increases competition in the short-term should not reduce incentives for competitors to invest in alternative facilities that will secure more competition in the long-term*"[57].

This means that NRAs need a long-term view of their decisions. Disruptive changes in prices of products or conditions belonging to the ladder of infrastructure would automatically lead to a distortion of competition and discrimination either against first movers or late entrants[58].

Only under consistent wholesale conditions, investments are allocated efficiently. If companies see any additional profit in climbing the ladder of investment, they will do so. Therefore, the ladder of investment shall also allow for possible migration processes from one rung of the ladder to the next[59].

Conclusion:	*A consistent pricing structure with regard to the ladder of investment is a prerequisite. To incentivise operators to climb the ladder, prices have to be set so that higher profit margins are possible for investments higher up on the ladder of investment. It is also necessary to implement effective migration rules.*

56 *ERG (2005a), exec. summary.*

57 *Recital 19 of Access Directive.*

58 *Ellig (2005).*

59 *On the aspect of static versus dynamic competition and the last mile problem involved see also Banerjee & Dippon (2006)*

With more competition law-based instead of sector-specific regulation, the consistent pricing within the ladder of investment may not be upheld. This will automatically cause some distortions within the ladder of investment. It has to be clearly stated that, under competition law, prices may include additional cost components compared to pure cost oriented prices and thus, can lead to significantly higher access prices.

Conclusion:	*The trend towards more competition law instead of sector-specific regulation could distort the consistency of prices within the ladder of investment. This, in turn, may have a significant economic impact on the business models of alternative operators.*

- ***Impact of new investments on new infrastructure (access holidays)***

Another issue is the topic of access holidays and (closely related) emerging markets. Following the idea of the 2002 framework operators investing in new infrastructure have argued in favour of granting access holidays for these investments (especially for FTTH or UMTS). The basis for this argument is laid down in recital no. 15 of the recommendation on relevant product and service markets[60].

The idea behind that argumentation is best described by Joseph Schumpeter[61], where the interplay of invention as creative destruction and imitation leads to more competition. The aim is to foster innovation and to allow these companies to have some first mover advantages. A more flexible approach than today's policy needs to be implemented. Otherwise, regulation will be too rigid to cope with the rapid technological changes of the future[62].

The EU commission has initiated a discussion of a revision of the current framework to become effective by 2009-2010[63]. It is clear that access holidays will reduce competitors' ability to offer products and services to consumers. This study clearly shows that this would have negative effects

60 *European Commission (2003), recital 15.*

61 *Schumpeter (1918).*

62 *For an overview see: http://europa.eu.int/information_society/policy/ecomm/tomorrow/ index_en.htm and on emerging markets*

http://europa.eu.int/information_society/policy/ecomm/doc/info_centre/public_consult/review/130706reviewpresentation.pdf, pp. 17-27.

63 *http://europa.eu.int/information_society/policy/ecomm/tomorrow/index_en.htm*

on prices and penetration rates, if neither infrastructure nor service-based competition is in place. In order to financially consider the high risk of investment undertaken, this is best achieved by allowing an appropriate return on investment.

***Conclusion**: Access holidays will reduce consumer benefits and have negative welfare effects. In order to create incentives for operators with significant market power, NRAs must consider the specific risks related to the investments in emerging markets in making their decisions.*

CONCLUSION

With the European Information society "i2010" initiative[64] Europe has started an ambitious project. A recent study has shown the positive impact of the European regulatory framework on the new member states[65]. The present study considers EU15 member states. The results describe that the different modes of infrastructure and service-based competition strategies harmonize as a growing number of countries enable competition on all rungs of the ladder of investment and market players diversify their market strategies. With this, prices tend to harmonize and Europe is stepwise developing towards a single market.

As infrastructure-based competition leads to higher innovation and penetration rates, NRAs should foster inter- and intra-modal infrastructure-based competition when applying the framework. With convergence and new technologies like broadband and VoIP emerging, access at the levels of networks, services, applications and devices and own infrastructure and thus ULL will become increasingly important for operators to be able to differentiate their products.

There are clear signs that infrastructure-based competition is more important to business customers and service-based competition is more important for residential customers. Therefore, if the majority of consumers are also to be able to benefit from competition then both liberalization strategies will have to be in place – in a balanced approach.

The study has also shown that infrastructure-based competition does have an immediate (downward) effect on prices, which tend to remain stable afterwards. Under such circumstances, less tariff rebalancing will occur and

64 *European Commission (2005).*

65 *Haberfehlner et al. (2006).*

alternative operators could refrain from investment in new infrastructure. On the other hand, service-based competition leads to significant tariff rebalancing in those parts of the value chain that are exposed to competition to a lesser extent. So these markets end up with higher access prices. In order to be competitive with infrastructure-based competition service-based competition needs much higher decreases in call prices to compensate for the increase in access costs.

On the other hand, if infrastructure-based competition alone is introduced, the risk of collusion remains in the sense that competitors only compete in the access market while increasing prices for other services. Hence both regulatory strategies are required. These results are consistent with the theory that price competition will take place in those parts of the value chain that are exposed to competition.

In most countries all rungs of the ladder of investment are now in place as regards the traditional network topologies. It seems that there is no pure and single way towards competition, but that markets need a healthy mixture of both, service-based and infrastructure-based competition. NRAs should consequently act in a stringent way regarding all rungs of the ladder in their pricing decisions. The market will react according to the strategic impact of the decisions made by NRAs, thus still relying on sector-specific regulation.

ADDENDUM

The results of this study with respect to infrastructure versus service-based competition rely on figures from the “old”, i.e. the circuit switched PSTN environment with data until 2005. Since the conclusion of our study new technologies have been introduced. This has led to a change of the markets especially with respect to broadband services.

The approach for serving broadband services relying most intensively on (own) infrastructure implying the requirement for the highest amount of investment is fiber to the home (FTTH), respectively fiber to the building (FTTB). Where operators go for this approach, it leads to a duplication of the network if such investment is undertaken by an ANO. ANOs may rely also still on regulatory decisions and wholesale products. For example, if alternative network try to mirror a fiber to the curb (FTTC) approach of the incumbent.

The Netherlands is a good example to see how a migration from a copper based network to a “all-IP” network is about to be implemented. Due to the fact that KPN dismantles their main distribution frames, carriers – which have worked on the basis of unbundled local loops today – have to make the choice of either moving up in the value chain (bit-stream) or down (FTTC). This means, they have to move into the direction of more or less investment.

Depending on how the carriers will react, we will also see a different degree of infrastructure versus service-based competition.

In the area of service-based competition, we additionally can see the developments with respect to resale. Resale is a vulnerable business model, due to the fact that the increase in bandwidth and the decrease in prices for end-users may put resellers at a certain risk in terms of price cost squeezes.

The demand from the customer side being based on the request for high bandwidths cannot necessarily be supplied by the existing wholesale products. Altogether, the demand for high bandwidths moves the market in total to the request for access products instead of conveyance products. This can be seen from the high increase in flat-rate products for Internet access as well as for call conveyance. Moving competition into the area of access competition therefore also implies that the access itself becomes more important. This is one of the reasons that trigger investments into access as the access line being the connection between the operator and the customer being more and more important.

The tendency that alternative operators conduct investments in terms of addressing the customers with their own infrastructure either by FTTC, FTTB or FTTH is therefore an interesting development. It seems to indicate (at least for some of the European countries) that competition is now moving really into infrastructure-based competition by a duplication of the last mile. It is, however, an open question whether this is rooted basically in the regulatory policy, a logical development of technology or a market driven necessity.

Market developments suggest that access competition is becoming increasingly more important. However, still a healthy mixture of service-based and infrastructure-based competition may remain important, especially for (rural) regions and markets where infrastructure-based competition is less likely.

REFERENCES

Banerjee A., Dippon C. (2006), "Communications Regulation and Policy Under Convergence: Advancing the State of the Debate", Paper presented at the 16th biannual ITS conference Beijing, June.

Bergman M. (2004), "Competition in services or infrastructure-based competition?", Swedish Competition Authority and Stockholm University, September.

Cave M. (2003), "Remedies for broadband services", www.itst.dk/static/. Konferencer%20og%20seminarer/EC-Experts%20Broadband%20_cave.pdf.

Cave M. (2004), "Making the ladder of investment operational", Paper presented to the European Commission, November.

Crandall R.W. (2005), "Competition and Chaos", Brookings Institution Press, 2005.

ECTA (2005), "ECTA Broadband Scorecard", Q2/2005.

Ellig J. (2005), "Costs and Consequences of Federal Telecommunications and Broadband Regulations", Working Paper in Regulatory Studies, George Mason University, February.

ERG (2005a), "Broadband market competition report", ERG (05) 23, May.

ERG (2005b), "Revised ERG Working paper on the SMP concept for the new regulatory framework", ERG (03) 09rev3, September.

European Commission (2000), "Commission Staff Working Document, "Europe's Liberalised Telecommunications Market - A Guide to the Rules of the Game", October 18th.

European Commission (2002), "Commission guidelines on market analysis and the assessment of significant market power under the Community regulatory framework for electronic communications networks and services" (2002/C 165/03), 11 July 11th.

European Commission (2003), "Recommendation on relevant product and service markets within the electronic communications sector susceptible to ex ante regulation" (C(2003) 497), 11 February 11th.

European Commission (2005), "Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions, "i2010 – A European Information Society for growth and employment", SEC(2005) 717.

European Parliament (2000), "Regulation no. 2887/2000 of the European Parliament and of the Council on Unbundled Access to the Local Loop".

European Parliament (2002), "Directive 2002/21/EC of the European Parliament and of the council of 7 March on a common regulatory framework for electronic communications networks and services" (Framework Directive), Official Gazette, no. L108/33, April 24th.

Flamm K. (2005), "An analysis of the determinants of broadband access", Telecommunications Policy Research Conference, October.

Haberfehlner K., Lundborg M., Pötzl J., Ruhle, E.O., Lichtenberger E. (2006), "Central and Eastern European countries way towards the Lisbon targets – ICT as driver for economic and social development", in Piepenbrock Schuster Consulting (ed.), "Telecommunications Markets in Central and South Eastern Europe - Market Developments and Regulatory Frameworks".

Henten A., Skouby K.E. (2005), "Regulation of local loop access", ITS Europe Conference, September 4-6th, Porto.

ITU (2001), "Italian Case Study", April 26th.

ITU (2001), "ITU regulatory implications of broadband workshop – ITU new initiatives programme", May 2-4th.

McNary R. (2001), "The Network Penetration Effects of Telecommunications Privatization and Competition", Stanford University.

Nicholls R. (2006), "Interconnection of next generation networks – a regulatory perspective", Paper presented at the 16th biannual ITS conference Beijing, June.

OECD (2003), "Developments in local loop unbundling", DSTI/ICCP/TISP(2002)5/ FINAL, 10th September 10th.

Piepenbrock Schuster Consulting (eds.) (2003), Anreize für Infrastrukturinvestitionen bei der Zusammenschaltung in der Telekommunikation.

Schumpeter J.A. (1918), The Theory of Economic Development: an Inquiry into Profits, Capital, Credit, Interest and the Business Cycle, Oxford Press, New York (1961).

Vogelsang I. (2002), "Theorie und Praxis des Resale-Prinzips in der amerikanischen Telekommunkationsregulierung", WIK Diskussionsbeitrag, no. 231, January.

Vogelsang I. (2005), "Resale und konsistente Entgeltregulierung", WIK Diskussionsbeitrag, no. 269, October.

Von Schewick B. (2007), "Towards an Economic Framework for Network Neutrality Regulation", Journal on Telecommunications and High Technology Law, Vol. 5, pp. 329-391.

ANNEX

Table 2. Decomposition of national analyses (1)

	Indicator 1		Indicator 2		Indicator 3	Indicator 4	Indicator 5		Conclusion
	Proportion of ULL and LS	Proportion of BSA and resale	Proportion of ULL and LS	Fixed broadband retail lines market shares of competitors	% -age infrastructure-based competitors of all competotors (incl. CATV)	% -age infrastructure-based competitors of all competotors (excl. CATV)	Completed introduction of indirect access in mid 2000 (1)	Introduction of LLU (0 = after 2001; 1 = before 2001; 2 = before 1999)	
A	3,4%	17%	3,4%	65%	85%	45%	100%	2	○
B	0,3%	21%	0,3%	49%	73%	3%	50%	1	--- (2)
D	5,3%	9%	5,3%	13%	72%	38%	50%	2	●
DK	5,5%	12%	5,5%	36%	83%	52%	100%	2	●
E	1,7%	25%	1,7%	44%	65%	16%	75% (3)	0	○
EL	0,1%	49%	0,1%	60%	53%	5%	0%	0	---
I	4,9%	15%	4,9%	30%	46%	46%	100%	1	○ (4)
IRL	0,1%	22%	0,1%	31%	35%	7%	100%	0	○
F	7,0%	23%	7.0%	52%	61%	56%	50%	1	● (5)
L	1,2%	10%	1,2%	26%	59%	25%	100%	0	---
NL	8,1%	0%	8,1%	56%	100%	100%	100%	2	●
P	0,9%	8%	0,9%	21%	92%	36%	50%	0	---
S	5,5%	14%	5,5%	60%	82%	52%	100%	1	●
SF	8,2%	10%	8,7%	36%	79%	68%	50%	1	●
UK	0,3%	65%	0,3%	54%	39%	2%	50%	1	○ (6)

○ = Service-based policy; ● = Infrastructure-based policy; --- = Indecisive

(1) The status is based on four categories of indirect access: carrier selection for local calls, carrier preselection for local calls, carrier selection for national calls, carrier preselection for national calls. Each category is weighted by 25%.

(2) Belgium, Greece, Luxembourg and Portugal were excluded as no allocation could be made.

(3) Completion for local calls in November 2000.

(4) The categorisation of Italy is due to the delay of the introduction of a standard offer for ULL and the early introduction of indirect access. In addition, the Italian NRA has focused on the introduction of a wholesale broadband product. Source: OECD (2003), pp. 22 and 45; ITU (2001), p.15.

(5) The categorisation of France is due to the aim of the NRA to foster infrastructure competition, as well as the success of line sharing. Source: OECD (2003), p.39; ERG (2005a).

(6) The results for UK are very significant. This may seem astonishing as the UK has fostered infrastructure competition in earlier years. These results are based on the situation in the new millennium, indicating that UK's regulatory policy has evolved from infrastructure-based to service-based competition.

Table 3. Decomposition of national analyses (2)

		Mean		Variance			
		Infrastructure based	Service based	Infrastructure based	Service based	P	Conclusions
Penetration rates	Penetration broadband	14,20%	8,10%	0,16%	0,08%	0,017	Significance
	Penetration DSL	10,40%	5,90%	0,02%	0,03%	0,003	Significance
	Fixed ISDN penetration 2004	14,20%	9,20%	0,40%	0,10%	0,154	
Prices 2004	National residential basket	28,36	31,21	17,96	15,96	0,283	
	National business basket	50,31	64,52	145,91	52,67	0,042	Significance
	Residential OECD composite basket	36,82	39,88	33,90	22,17	0,360	
	Business OECD composite basket	74,32	94,90	240,29	188,78	0,045	Significance
	Incumbent's basic monthly PSTN rental charge for business customers	11,88	18,27	4,42	2,32	0,000	Significance
	Incumbent's basic monthly PSTN rental charge for residential customers	13,49	17,79	5,92	6,81	0,022	Significance
Price changes between 2000 and 2004	National residential basket	6,10%	-6,60%	0,20%	0,90%	0,036	Significance
	National business basket	3,20%	-9,40%	0,40%	1,00%	0,049	Significance
	Residential OECD composite basket	-1,10%	-16,30%	0,50%	0,60%	0,009	Significance
	Business OECD composite basket	-3,80%	-11,90%	0,50%	1,80%	0,275	
	Incumbent's basic monthly PSTN rental charge for business customers	3,90%	27,40%	1,60%	4,10%	0,062	
	Incumbent's basic monthly PSTN rental charge for residential customers	8,80%	28,60%	2,00%	5,40%	0,146	
Prices 2005	Residential monthly rental	14,90	17,14	5,20	15,84	0,307	
	Business monthly rental	12,92	17,54	2,94	9,84	0,026	Significance
	Average monthly expenditure (composite basket), business	75,83	86,80	139,77	295,70	0,267	
	Average monthly expenditure (composite basket), residential	38,83	38,20	26,97	22,70	0,838	
Prices 2004 adjusted for PPP	National residential basket	26,75	32,02	28,31	44,26	0,192	
	National business basket	47,69	66,34	190,50	157,71	0,044	Significance
	Residential OECD composite basket	34,71	41,01	49,70	74,58	0,229	
	Business OECD composite basket	70,34	97,35	331,29	344,79	0,040	Significance
	Incumbent's basic monthly PSTN rental charge for business customers	11,27	18,61	8,11	3,75	0,001	Significance
	Incumbent's basic monthly PSTN rental charge for residential customers	12,76	18,06	9,95	5,58	0,011	Significance

Broadband Investment and the Threat of Regulation: Preventing Monopoly Exploitation or Infrastructure Construction?[66]

ULRICH BLUM
President of the Halle Institute for Economic Research

CHRISTIAN GROWITSCH
Head of Department at the Scientific Institute for Infrastructure and Communication Services WIK

NIELS KRAP
Halle Institute for Economic Research

***Abstract**: Recently, investments in new generation networks in Germany have been curtailed, since potential investors required this new network be unregulated. To develop a regulator's strategy that allows investments to occur but prevents monopolistic prices, we model an investor's decision problem under a threat of regulation. We show that the mere threat of a regulator's intervention may prevent supernormal profits without actual price regulation. The regulator, on the other hand, can influence both the investment decision and the investor's price via her signals on regulation probability and price.*

INTRODUCTION

In summer 2005, Deutsche Telekom announced its plans to build a new generation broadband fiber optics network. The initial investment was said to be around 3 billion Euros. However, Deutsche Telekom decided as a precondition for this new network not to be regulated with respect to pricing and third party access.

German regulation authorities announced their refusal to concede to Telekom's pressure. They suggested that Telekom and its competitors, mostly service providers that rent capacity from the dominant incumbent, agree on rules as how to manage access to the new infrastructure. Following this, Telekom let it be known that they were unwilling to share the new capacities with competitors, claiming that competitors should undertake the irreversible network investments (sunk costs) by themselves. Unless the new technologically-leading infrastructure was exempted from regulation,

66 *Partial reprint from Blum U., Growitsch C. & Niels Krap (2007), "Broadband Investment and the Threat of Regulation: Preventing Monopoly Exploitation or Infrastructure Construction?," Review of Network Economics, Vol. 6(3), pages 342-354, September with the permission of CRA International.*

Telekom threatened the investment to be made in other areas or countries. Hence, the German regulator faced a difficult situation: allowing Telekom to have its way may endanger future competition on next generation networks in telecommunication in Germany. However, if the regulator decided access regulation to hold, the infrastructure would not be set up.

In summary, the regulatory agency faced an issue of first and second degree errors: if she regulated an area which – from the economic point of view – should be left unregulated, such an over-regulation may foreclose welfare enhancing investments. If, however, she did not regulate an area which required regulation, under-regulation could inhibit competition and facilitate market power exploitation.

In this paper, we analyze investment decisions on new generation networks. Therefore, we discuss a firm's decision problem under a threat of regulation from a game-theoretic perspective. The decision whether to invest or not depends on the probability of regulation and its assumed impact on investment returns. Depending on the investor's expectation on these parameters, he will decide whether the investment is favorable or not, and which price is optimal. This price can be expected to be lower than a non-regulated profit maximizing price, since the potential investor presumably tries to circumvent regulation and reduce the probability of intervention, respectively. Thus, the mere threat of a regulator's intervention may prevent supernormal profits without actual price regulation. The regulator, on the other hand, can influence both investment decision and the investor's price via his signals on regulation probability and price. These signals can be considered optimal if they simultaneously allow investment and minimize the investor's price. Accordingly, wrong signals by the regulator may prevent investments. Hence, we model an investment decision under uncertainty (of regulation) to develop a welfare maximizing regulation strategy.

LITTERATURE

Previous research on the relationship between investment and regulation has discussed either impacts of specific regulatory regimes or incentives of underinvestment due to policy uncertainty. Continuing research on dynamic efficiency issues of regulation discussed by Mandy & Sharkey (2003) and Littlechild (2003), a current work by Evans & Guthrie (2005) addresses the negative incentives on investment imposed by total element long run incremental cost (TELRIC) regulation, and finds that within such a framework a capital asset pricing model application identifies an allowed risk premium to be crucial for sustainable investments. Indeed, Evans and Guthrie's models give interesting insights into the investment incentives of specific regulatory regimes. Unfortunately, their models assume a universal service obligation as well as a general revenue regulation.

A second strand of literature discusses policy uncertainty or – more specifically – regulators' ex-post opportunism (potential hold-up). Recent work on different network industries, for example Ishi & Yan (2004), Saphores et al. (2004) and Dobbs (2004) confirms the hypothesis of delayed infrastructure investments as addressed by Teisberg (1993), who showed that rational firms might delay investment when facing uncertain or asymmetric profit and loss restrictions. However, previous research on investment under regulation has to a certain extend neglected welfare enhancing aspects of regulatory uncertainty, the issue of regulatory threats, which is basically the threat of governmental intervention in case of inadequate price levels[67]. The political intention is the firms restricting their prices voluntarily (the so called light-hand regulation approach, for an overview on network industries see Haucap et al. 2006). Developed by Glazer & McMillan (1992), there have been numerous applications on different network sectors (for an example of the British airport sector see Starkie 2001, and Acutt & Elliott 2001 for the experiences of the UK electricity generation industry). Brunekreeft (2004) translates the idea of regulatory threat to the threat of ex-post antitrust intervention, finding that under certain conditions the latter can work in similar fashion and also induce a voluntary price cap. The work on regulation by threat of intervention has neglected, as yet, to emphasize its relevance for investment decisions in network sectors[68]. This paper fills that void in previous research and applies the concept of light hand regulation on new infrastructure investment. It contributes to the current discussion of regulatory options for next generation networks and regulation's effects on dynamic efficiency and innovation (see, for example Baake et al. 2005), since it provides a feasible solution to the trade-off between static or allocative and dynamic efficiency or innovation/technological change.

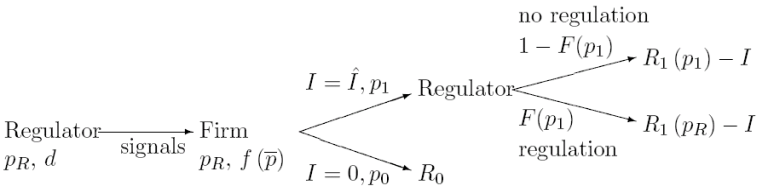
67 *Admittedly, Sappington (1986) deal with regulatory uncertainty, whereas from a different perspective. He analyses the positive effects of information asymmetries on firms' investment, not the idea of putting a company under regulatory risk, as we do.*

68 *There has been some research on a phenomenon called trigger price regulation, which is somewhat comparable to our approach. In a seminal article, Salant & Woroch (1992) introduce the phrase and show, given perfect information, the regulator can allow for new investments while not directly regulating. A credible commitment of not exploiting producer's rents is however very sensitive to the symmetric information assumption, which can be considered as rather strong.*

THE MODEL

Applying a game theoretic model, we analyze a firm's new generation network investment decision facing a possible regulatory intervention. We assume the regulator to be benevolent but imperfectly informed about the investor's cost structure.

Figure 2. Decision tree of the game-theoretical regulation model



This figure shows the options of the market agents and the sequence of decision in the investment game. In a first step, the regulator signals her perception of a cost-covering price – the regulation price (p_R) – and about her tolerance (d) on the firm's pricing decision. Both can be interpreted as a probability of intervention function ($F(p_1)$) depending on the firm's price. The regulator's true tolerance limit – the so-called intervention price (p_R+d) – can be concealed to the firm, exposing the investor to regulatory risk. Thus, signaling the tolerance enables the regulator to – at least partially – uncover the true costs of the firm.

If the regulation authority signaled the intervention price perfectly, the investor would set exactly that price or – in the case of an unfavourably low expected intervention price – refrain from investment. If the firm invests ($I=\hat{I}$), the regulator decides whether the firm's price (p_1) requires an intervention. Under regulation, which has the probability $F(p_1)$, the regulator intervenes and permits the firm to set only the regulation price p_R. Thus, the firm obtains a profit of $R(p_R)$- $\hat{I}$. If the price set by the firm lies within the regulator's tolerance, it earns a profit of $R(p_1)$- $\hat{I}$.

If the investor's expectation of the regulation price is below a profitable level, it will refrain from investment, earning a profit of R_0. Such a situation is economically undesirable and should be avoided by the regulator. But due to her lack of information it is possible, that the regulator overestimates the chances and earnings and underestimates the risks and costs of the investment and sets the intervention price too low.

Without perfect announcement (an imperfect signal) the firm can, if it invests,

- either set a price above the intervention price, so that the regulator intervenes and sets the regulation price or
- set a price below the intervention price due to an effective regulatory threat.

In both cases, the resulting market price is lower than or equal to the intervention price and therefore welfare-increasing.

COMMENTS

Major result of this model is that the optimal price of an investor facing the threat of regulation is lower than the expected intervention price. Intervention or actual regulation becomes unnecessary; the market price can be expected to be (significantly) lower than an unregulated monopoly price (Cournot-price). Obviously, the firm's optimal price (and thus its profit) in this model rises with increasing regulation price and increasing tolerance.

Sappington (1986) shows that given a perfectly informed regulator, firms will have no incentive to invest in efficiency enhancing investments since the resulting cost-savings would be completely transferred to consumers. Therefore, they suggest that the regulator may abstain from better information. As we consider a regulator with long-term orientation as introduced by Salant & Woroch (1992), these results do not directly contradict our analysis. The regulator in our model allows for positive expected returns to provide an incentive for new infrastructure installation rather than holding up the firm.

In such a setting, actual regulation becomes unnecessary. The mere threat of regulation prevents monopolistic prices while it allows profitable investment – if her signals are not to restrictive and therefore foreclosing.

On the one hand these results show that with rising uncertainty about the actual investment costs and market conditions a regulation authority with the goal of high investment activity should give either a signal of a higher regulation price or a higher tolerance to the investment in relation to the deviation from that price. On the other hand, the more the regulator knows the tolerance can sink and thereby set the intervention price nearer to the price at which the investment just amortizes. According to this model a convincing and purposeful regulation threat can replace an implemented regulation. However, it remains to be noted that in most cases the price set by the investor does not correspond to the economically optimal price. Finally there is – similar to the situation with patents and the protection of innovations – a political and economical tradeoff between the investment with more utilization of market power and non-investment. In addition should be mentioned that there is also the possibility that the regulation threat

prevents the investment if the regulation authority underrates the necessary investments and/or overrates the future profits and options, and thus signals a too small tolerance and/or sets a non-cost-cutting regulation price.

CONCLUSION

Recapitulating our findings, we state that as long as the regulator is uncertain about cost and demand structure in the market of the infrastructure to be enhanced, she should not be acting too intolerantly, since such behaviour may prevent a welfare increasing investment. These results are consistent with previous research on the effectiveness of regulatory threats in particular. Concerning the case of Deutsche Telekom, the German regulator should – ex ante – leave the infrastructure investment unregulated and signal the regulation price. This should encourage Deutsche Telekom to invest while preventing it from exploiting its monopolistic power. From a dynamic perspective, such a light-handed regulation may encourage additional – and competitive – infrastructure investment, increase technological development and economic welfare, and make regulation redundant in the future. Therefore, our findings show that the concept of regulating by the threat of intervention is not only applicable to existing infrastructure but also to new investments as well.

REFERENCES

Acutt M., Elliott C. (2001), "Threat-based Regulation and Endogenously Determined Punishments", Lancaster University Management School Working Paper 2001/007.

Baake P., Kamecke U., Wey, C. (2005), "A Regulatory Framework for New and Emerging Markets", Communications and Strategy, Vol. 60, pp 123–136.

Blum U., Growitsch C., Krap N. (2007), "Broadband Investment and the Threat of Regulation: Preventing Monopoly Exploitation or Infrastructure Construction?", Review of Network Economics, Vol. 6, No. 3, pp 342–354.

Brunekreeft G. (2004), "Regulatory Threat in Vertically Related Markets: The Case of German Electricity", European Journal of Law and Economics, Vol. 17, pp 285–305.

Dobbs I.M. (2004), "Intertemporal Price Cap Regulation under Uncertainty", The Economic Journal, Vol. 114, pp 421–440.

Evans L.T., Guthrie G.A. (2005), "Risk, Price Regulation, and Irreversible Investment", International Journal of Industrial Organization, Vol. 23, pp 109–128.

Glazer A., McMillan H. (1992), "Pricing by the Firm under Regulatory Threat", Quarterly Journal of Economics, Vol. 107, pp 1089–1099.

Haucap J., Heimeshoff U., Uhde A. (2006), "Credible Threats as an Instrument of Regulation for Network Industries", in Welfens P.J.J. (ed.), "Regulatory Changes, Innovations and Investment Dynamics", Springer.

Ishi J., Yan J. (2004), "Investment under Regulatory Uncertainty: US Electricity Generation Investment Since 1996", Working Paper CSEMWP-127, Center for the Study of Energy Markets, University of California.

Littlechild S. (2003), "Reflections on incentive regulation", Review of Network Economics, Vol. 2, pp 289–315.

Mandy D.M., Sharkey W.W. (2003), "Dynamic Pricing and Investment from Static Proxy Models", Review of Network Economics, Vol. 2, pp 403–439.

Salant D.J., Woroch G.A. (1992), "Trigger Price Regulation", RAND Journal of Economics, Vol. 23, pp 29–51.

Saphores J.D., Gravel E., Bernard J.T. (2004), "Regulation and Investment under Uncertainty: An Application to Power Grid Interconnection", Journal of Regulatory Economics, Vol. 25, pp 169–186.

Sappington D.E.M. (1986), "Commitment to Regulatory Bureaucracy", Information Economics and Policy, Vol. 2, pp 243–258.

Starkie D. (2001), "A New Deal for Airports?", in Robinson C. (ed.), "Regulating Utilities: New Issues, New Solutions, Cheltenham", Elgar Publisher.

Teisberg E.O. (1993), "Capital Investment Strategies under Uncertain Regulation", The RAND Journal of Economics, Vol. 24, pp 591–604.

Service-Based versus Infrastructure-Based Competition: A Real Options Approach[69]

KEIICHI HORI
Faculty of Economics, Ritsumeikan University

KEIZO MIZUNO
School of Business Administration, Kwansei Gakuin University

Abstract: *This paper compares two specific competition schemes – service-based and infrastructure-based competition – by focusing on a firm's incentive to invest in network infrastructure. We show that when monopoly rent is large, infrastructure-based competition means the initial introduction of infrastructure is undertaken earlier than under service-based competition. However, when both monopoly rent and the degree of uncertainty are small, service-based competition brings about the earlier initial introduction of infrastructure than under infrastructure-based competition. The paper includes discussion of the policy implications of these findings.*

INTRODUCTION

Infrastructure investment is crucial to economic welfare in the long run. This paper addresses how infrastructure investment in telecommunication industries can be promoted. In particular, we examine the effect of the choice of competition scheme on a firm's incentive to invest in network infrastructure. To this end, we focus on two specific competition schemes: service-based competition and infrastructure-based competition. In service-based competition, entrants can enter the market by accessing an incumbent's network infrastructure when desired. On the other hand, infrastructure-based competition requires entrants to construct their own infrastructure in order to enter the market. In this paper, we clarify the conditions under which one competition scheme would induce a firm to invest earlier in infrastructure than the other.

69 *This paper is based on Hori and Mizuno (2007).*

RELATED FINDINGS IN THE LITERATURE

Several studies have compared competition schemes in an open access environment. A closely related analysis to this paper is Bourreau & Doğan (2005). By focusing on an incumbent's incentive for unbundling and the incentive to set an access charge, Bourreau & Doğan (2005) show that an incumbent has an incentive to set too low an access charge. As a result, the entrant builds its own infrastructure too late from a welfare viewpoint. Empirical research by Kaserman & Ulrich (2002) shows the effects of infrastructure-based entry versus resale entry on competition. According to their results, resale entry seemingly has a more drastic effect on competition than infrastructure-based entry. This is because resale entry reduces the incumbents' shares in the long-distance telecommunication market more than with infrastructure-based entry. Guthrie (2006) also provides valuable discussion about the two competition schemes, with an emphasis on their effects on the firm's timing of investment.

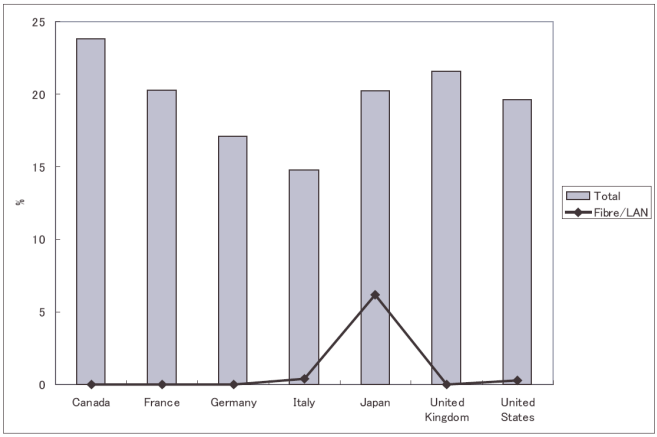
DEPARTURE FROM PREVIOUS LITERATURE

Our study differs from previous work in two main ways. First, most studies on open access policy, with the exception of Biglaiser & Riordan (2000) and Pindyck (2005), do not deal with uncertainty and irreversibility, whereas we examine the firm's incentive for irreversible investment under uncertainty. Uncertainty and irreversibility are decisive for firms in network industries, such as telecommunications. Those involved in network industries, including researchers and policy makers, commonly recognize both of these elements. For example, Guthrie (2006) argues that the dynamics, investment irreversibility, and uncertainty are all essential features of regulated competition.

Indeed, the net present value rule, which states that an investment should be undertaken only if its net present value is positive, is inappropriate for firms operating under uncertainty and irreversibility. This is why we employ a real options approach to discuss this issue. The real options approach applies option concepts to the valuation of real assets under uncertainty. It has become an important growth area in investment theory (see Dixit & Pindyck 1994) for a basic treatment of the tools employed). Indeed, Alleman & Noam (1999) suggest the application of the real options approach to the telecommunications industry. We then examine how uncertainty affects the priority of the two competition schemes in terms of a firm's incentive to invest in telecom infrastructure.

A second major difference in our work is that while previous studies assume that firms' positions in a market are exogenous, in our model they are endogenously determined. This is because we would like to consider the firm's preemption incentive or first-mover advantage (see Fudenberg & Tirole 1985, Katz & Shapiro 1987). In fact, telecommunication industries have given birth to many situations where the preemption incentive is appropriate.

Figure 3. The Penetration Rates of Broadband Services



Source: OECD Broadband Statistics, December 2006.

The previous figure shows the penetration rates of broadband services, including fiber-optic networks, in G7 countries as of December 2006. One trait that separates Japan from the other countries is the high penetration of fiber-optic or local area network (LAN) services. In Japan, 30% of subscribers of broadband services utilize fiber-optic networks, compared to relatively few subscribers in the remaining G7. This reminds us that the Japanese Telecom regulatory authority employed some policies that stimulated investment in fiber-optic networks. Indeed, the e-Japan plan was set in place to ensure that broadband networks were accessible to 30 million households by 2001. This plan (and the following related plans) explicitly states that the broadband networks established throughout Japan should be developed with competition among the firms. Since 2001, several players—including NTT East and West, K-Opticom, Yusen etc.—have built their own fiber-optic cable networks to obtain a first-mover advantage. On the other hand, it is argued that the US is not necessarily a well-developed country in the field of fiber-optic networks. As a result, in 2003 the Federal Communications Commission (FCC) adopted new rules regarding the network unbundling obligations of incumbent local phone carriers, with the aim of providing incentives for carriers to invest in broadband infrastructure.

Investment in mobile networks to enhance a firm's service coverage is another example that relates to the issue of preemption among firms. In mobile network markets, mandated access or mandated roaming is a policy issue. Interestingly, regulatory agencies have taken different decisions on this issue in various countries (see Hausman (2002) for a discussion). Fully providing several types of infrastructure in developing countries is also appropriate for preemption, especially when privatization and unbundling are simultaneous. In telecommunications, the R&D market with licensing is a good example that fits the issue of preemption.

In sum, these examples suggest the importance of the study of a firm's preemption incentive in an open access environment. We should note that allowing access to a follower (i.e., an open access environment) gives a follower an advantage in the sense that the follower can avoid the sunk cost of infrastructure by accessing a leader's infrastructure with payment of an access charge. This follower's advantage may weaken a firm's preemption incentive (i.e., it may deter the introduction of new network infrastructure in a market). By analyzing the effects of allowing access on the preemption incentive in an open access environment, we examine the priority of the two competition schemes in terms of the rapidity of the initial construction of network infrastructure.

A THEORETICAL MODEL

To examine this issue, Hori and Mizuno (2007) introduced a two-firm model with stochastically growing demand in an open access environment. No firm establishes its facility at the beginning. Each firm requires two types of facilities to serve consumers in the market: a production facility and a network (infrastructure) facility. Investments in these facilities are irreversible. Under service-based competition, a follower without a network can utilize the existing network for production by paying a usage access charge. On the other hand, under infrastructure-based competition, a follower needs to invest not only in a production facility, but also in a network.

The demand for goods produced and the profit flows of the firms are assumed to grow stochastically. The building of a bypass (i.e., an additional network supply) is assumed to improve the quality of goods or cause a positive externality: this is reflected in an increase in each firm's profit flow.

Taking a competition scheme and the level of access charge as given, each firm decides when to invest. When entering the market, a firm decides the output level, and the product market clears. When examining service-based competition in a stochastically demand-growing environment, it is natural to focus on the situation where a follower first enters with access to a leader's

network, and in the future, builds a bypass itself. The equilibriums under the two competition schemes are then derived, and we can compare the outcomes.

COMPARISON OF THE TWO COMPETITION SCHEMES

Let us report the results derived from the analysis of the theoretical model.

Result 1:	*Under service-based competition, a follower enters the market earlier and builds a bypass later than under infrastructure-based competition.*

The latter result is explained by the replacement effect: a follower that already gains profit through access to the leader's network has less incentive to build a bypass than a potential follower that now intends to enter the market. The replacement effect also appears in Bourreau & Doğan (2005). In addition, we can verify that as the degree of uncertainty increases, the difference in a follower's investment timing becomes greater between the two competition schemes.

Next, let us compare a leader's entry timing for the two competition schemes. Three elements (i.e., monopoly rent, the level of access charge, and the degree of uncertainty) are key factors in determining the priority of competition schemes in terms of a leader's investment timing.

Result 2:	*When monopoly rent is sufficiently large relative to duopoly rent, infrastructure-based competition induces a leader to invest earlier than under service-based competition.*

The intuitive reasoning is simple. Since a follower under infrastructure-based competition enters the market later than one under service-based competition, a leader under infrastructure-based competition enjoys monopoly rents for a longer period than under service-based competition. Hence, the preemption incentive of a leader under infrastructure-based competition is larger than under service-based competition.

This finding suggests a relationship between the investment timing of infrastructure and the degree of competitiveness in a product market. In particular, when product market competition is severe, monopoly rent is sufficiently large relative to duopoly rent. In this case, infrastructure-based competition induces a leader to invest in infrastructure earlier than under service-based competition.

We next examine the effect of a change in the level of access charge on investment decisions in the two competition schemes. Concerning the follower's investment decision, a follower's entry under service-based competition is earlier than under infrastructure-based competition, with a bypass being built later. However, as the access charge increases, a follower's entry under service-based competition is later.

The effect of the access charge on a leader's entry timing under service-based competition was examined in Hori & Mizuno (2006). According to the finding therein, an increase in access charge induces a leader to enter the market early. That is, the incentive for preemption is enhanced by an increase in access charge under service-based competition. However, it is difficult to analytically derive a necessary and sufficient condition for service-based competition to make a leader's entry earlier than infrastructure-based competition and vice versa because of nonlinearity in the effect. Hence, we only report the sufficient condition for service-based competition. In fact, when both monopoly rent and uncertainty are small, the preemption effect under service-based competition is stronger than under infrastructure-based competition.

Result 3:	*When both monopoly rent and uncertainty are small, a leader under service-based competition enters the market earlier than under infrastructure-based competition.*

Result 3 is explained as follows. Remember that a follower's entry under service-based competition is earlier than under infrastructure-based competition. This implies that a leader under infrastructure-based competition can earn a monopoly rent for a longer period. As stated earlier, however, a decrease in uncertainty makes a follower's entry timing under service-based competition closer to that under infrastructure-based competition. That is, the difference in the period in which the leader can earn a monopoly rent between infrastructure-based competition and service-based competition becomes shorter with the reduction in uncertainty.

Moreover, for the period in which a follower accesses a leader's network, the leader under service-based competition can obtain access profits. The access profit can then compensate for the loss caused by a shorter period of monopoly rent under service-based competition. Therefore, a leader under service-based competition has a greater incentive to enter than a leader under infrastructure-based competition, when both the monopoly rent and uncertainty are small.

CONCLUSION: SOME POLICY IMPLICATIONS

Taking the competition scheme and level of access charge as given, we examined the relationship between the choice of competition scheme and the investment decisions of both leaders and followers. Although all of the results concern which particular competition scheme can achieve earlier investment timing, earlier timing is not necessarily better in terms of welfare. In fact, it is difficult to determine which of the two competition schemes is better from a welfare viewpoint.

Nevertheless, we can obtain some policy implications for the choice of competition scheme. According to Result 1, a follower under service-based competition is more proactive in entering a market because it need not incur the large sunk costs of building its own network. In this respect, service-based competition realizes a competitive environment in the product market earlier than infrastructure-based competition. In turn, this contributes to the enhancement of economic welfare. This is consistent with a crucial aspect of service-based competition pointed out in the literature. That is, a competitive environment can be achieved earlier under service-based competition because of non-duplication of the infrastructure investment's sunk cost. In fact, this provides a partial rationale for the unbundling provisions of the 1996 Telecommunications Act.

However, service-based competition simultaneously entails the late construction of a bypass. This implies the later introduction of the positive externality generated by the prevalence of infrastructure. Because consumers also obtain benefit from the positive externality, service-based competition may be harmful. In sum, service-based competition includes benefits and costs from a welfare viewpoint when focusing on the follower's entry. Note also that the magnitude of the benefits and costs becomes larger as uncertainty increases.

On the other hand, according to Results 2 & 3, service-based competition may deter the initial construction of network infrastructure that depends on the environment surrounding firms. Therefore, we cannot confirm whether service-based competition enhances social welfare with reference to a leader's entry.

In reality, the negative effect of service-based competition on social welfare already appears to be recognized, particularly in telecommunications. For example, it is argued that the US is not necessarily a well-developed country in the field of fiber-optic cable networks. Several aspects of the regulatory regime in the 1996 Telecommunications Act, such as asymmetric regulation between cable TV companies and incumbent local exchange companies (ILEC), are considered as elements that deter investment in broadband networks (see Crandall & Alleman (2003)).

We can explain the non-development of fiber-optic cable networks by applying our analytical results. According to our findings, this is because the monopoly rent has been large in the US telecommunication industry. In fact, because cable TV networks were prevalent in the US broadband market, the advantage of the leader introducing a fiber optic cable network with a higher-speed Internet service is larger. In addition, the demand for broadband services has not been readily estimated (i.e., demand is uncertain), because wireless and the Internet have changed the way people communicate. Hence, in this situation infrastructure-based competition has been superior to service-based competition in terms of early investment in new infrastructure. As a result, in 2003 the Federal Communications Commission (FCC) adopted new rules regarding the network unbundling obligations of incumbent local phone carriers, with the aim of providing incentives for carriers to invest in broadband (i.e., a change in the competition scheme towards infrastructure-based competition).

REFERENCES

Alleman J., Noam E. (eds.) (1999), "The New Investment Theory of Real Options and Its Implications for Telecommunications Economics", Boston, Kluwer Academic Publishers.

Biglaiser G., Riordan M. (2000), "Dynamics of Price Regulation", Rand Journal of Economics, Vol. 31, 744–767.

Bourreau M., Doğan P. (2005), "Unbundling the Local Loop", European Economic Review, Vol. 49, 173–199.

Crandall R., Alleman J. (eds.) (2003), "Broadband: Should We Regulate High-Speed Internet Access?", Washington DC, The Brookings Institution.

Dixit A.K., Pindyck R. S. (1994), "Investment under Uncertainty", Princeton NJ, Princeton University Press.

Fudenberg G., Tirole J. (1985), "Preemption and Rent Equalisation in the Adoption of New Technology", Review of Economic Studies, Vol. 52, pp. 383–401.

Guthrie G. (2006), "Regulating Infrastructure: The Impact on Risk and Investment", Journal of Economic Literature, Vol. 44, pp. 925–972.

Hausman J. (2002), "Mobile Telephone", in Cave M.E., Majumdar S.K. and Vogelsang I. (eds.), Handbook of Telecommunications Economics, Vol. 1, Amsterdam, Elsevier Science B.V., pp. 563–604.

Hor, K., Mizuno, K. (2006), "Access Pricing and Investment with Stochastically Growing Demand", International Journal of Industrial Organization, Vol. 24, pp. 795–808.

Hori K., Mizuno K. (2007), "Competition Schemes and Investment in Network Infrastructure under Uncertainty", Paper presented at the 6th Conference on Applied Infrastructure Research, Berlin, October 5-6.

Kaserman D.L., Ulrich, M. (2002), "The Competitive Effects of Resale versus Facility-Based Entry: Evidence from the Long-Distance Market. Telecommunications Policy", Vol. 26, pp. 415–424.

Katz M.K., Shapiro C. (1987), "R&D Rivalry with Licensing or Limitation", American Economic Review, Vol. 77, pp. 402–429.

Pindyck R.S. (2005), "Pricing Capital under Mandatory Unbundling and Facilities Sharing", NBER Working Paper No. 11225.

Part II
Focusing on Mobile Issues

The Ladder of Investment in the Mobile Market[70]

MARC BOURREAU
Institut Telecom, Telecom ParisTech, Department of Economics and Social Sciences, and CREST-LEI

JOEFFREY DROUARD
Institut Telecom, Telecom ParisTech, Department of Economics and Social Sciences

LAURENT FERRALI
Université Paris 1 Panthéon Sorbonne

Abstract: *The "ladder of investment" (Cave 2006) is often cited as a relevant regulatory mechanism to stimulate facility-based competition in the fixed broadband market, but is very rarely mentioned as a means to facilitate the deployment of alternative infrastructures in the mobile market. In this paper, we examine the relevancy of this concept for the mobile industry, and we discuss whether the concept could be applied to encourage the development of alternative mobile facilities.*

INTRODUCTION

In the telecommunications industry, entrants rarely have enough financial resources to deploy from the outset network infrastructures, which cover an entire territory and/or the totality of the value chain. Consequently, entrants tend to be gradual in both the timeframe in which they invest in the network elements necessary to provide telecommunications services, as well as the geographical areas across which they deploy their networks. To minimize this transitional phase and, hence, to promote investment in alternative infrastructures, Martin Cave (2006) has suggested, for the fixed broadband market, to classify the different network segments according to their level of replicability and to specifically regulate each segment. This doctrine is often referred to as the "ladder of investment".

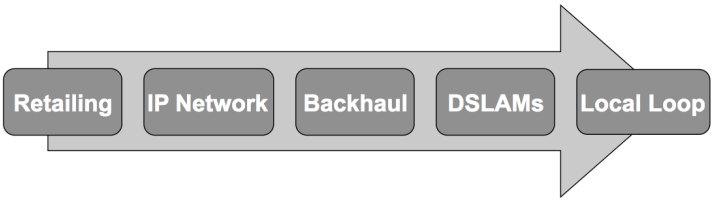
Implementing the ladder of investment requires two steps. First, to identify the replicable network segments. Second, to classify these segments according to their degree of replicability. In particular, a segment of the infrastructure is considered difficult to replicate if economies of scale, scope or density are important. Notice however that economies of scale or scope

70 *The authors thank Laurent Benzoni and Gérard Pogorel for their useful remarks and Sara Clignet for her editorial assistance.*

constitute a barrier to entry only when investments are irreversible and the environment is uncertain.

To elaborate the following ladder of investment for the fixed broadband market, Martin Cave uses an historical approach[71], and considerations on the costs of network elements.

Figure 4. The ladder of investment in fixed broadband networks



Source: From Cave (2006)

Once the ladder has been determined, applying the "ladder of investment" doctrine requires identifying the positions of network operators on this ladder and elaborating a pricing mechanism, to induce entrants to climb the ladder (either based on rising access charges or on the announcement of the withdrawal of mandatory access after a given date).

This article aims to analyze whether the concept of the ladder of investment, which is widely viewed as a relevant concept for the fixed broadband market, can be applied to the mobile market as well. To this end, we shall first construct a “ladder” for the mobile market and then, second, discuss the relevancy of this ladder.

In the first section, we will begin by constructing a ladder of investment for the mobile market. We will develop our analysis, putting aside the technological and economic limits due to the scarcity of spectrum. We will only focus on identifying an ad hoc classification of the degree of replicability for the different mobile network elements.

Two factors have encouraged regulatory authorities to apply the concept of the ladder of investment in the fixed broadband market: first, to help new entrants build gradually a customer base and, second, to help them acquire revenues progressively. In the second section, we will discuss whether the gradual acquisition of a customer base and of revenues is also essential for

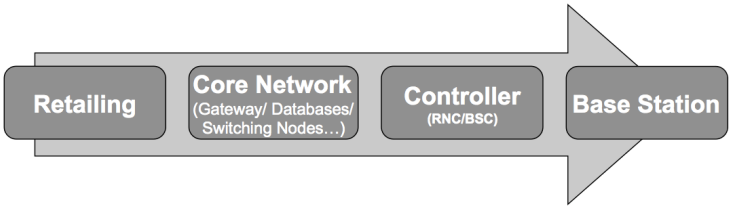
71 *Cave adopts an empirical approach and looks at the degree at which operators have built out competing networks.*

the development of facility-based competition in the mobile market. The discussion is followed by an analysis of available regulatory tools to help network operators climb this ladder. The final part focuses on mobile market specificities.

A LADDER OF INVESTMENT FOR THE MOBILE MARKET[72]

In this section, we ignore the scarcity of spectrum and its implications, to be discussed later. Under this assumption, using the degree of replicability as a criterion for classification, we propose the following ladder of investment for the mobile market:

Figure 5. A ladder of investment for the mobile market[73]



Each ladder rung is detailed below.

- ***Base Stations***

Base stations are the least easily replicable segment of the mobile network for two reasons. First, because of their high costs due to the large number of

72 *The data provided in this section are from the report of the European Commission (June 25, 2002).*

73 *Due to the constraints on access to radio spectrum, operators which do not own a spectrum license can only invest in the retailing and the core network. Operators without a license are called mobile virtual network operators (MVNO). There are three types of MVNOs, based on their level in the ladder of investment. Classic Service Providers merely resell minutes, Enhanced Service Provider own their own SIM cards and some network elements, whereas Full MVNOs own a full mobile network, except for the wireless local loop elements.*

stations required for full territorial coverage[74]. Second, because of the economies of density associated with this investment.

Base station costs are predominantly fixed and characterized by significant economies of density[75]; meaning that profitability depends on the size of the potential customer base. These economies of density could be viewed as a barrier to entry. However, a significant proportion of these investments is not specific to the technology used (for example, the facilities[76]) and should therefore not be considered as sunk costs.

- ***Station Controllers (RNC / BSC)***[77]

During initial deployment, and when capacity constraints are not binding, a controller can monitor up to approximately 400 base stations. As the network grows, the number of base stations that a controller can monitor decreases. Once the network is fully deployed, a controller manages between 25 and 50 base stations, depending on the area's urbanism. Deployment costs represent about 20% of the total cost of the wireless local loop.

It is important to note that controllers are not specific to the mobile technology used, and that the RNC and the BSC are interoperable. Interoperability, therefore, implies lower costs in the transition from the GSM network to the UMTS network.

74 *The fixed broadband network's DSLAM can be compared to the mobile network's base stations. French mobile operators each own between 10,000 and 15,000 base stations. For the fixed broadband market, a network operator has to install DSLAMs in each of the 12000 main distribution frames (MDF) to cover the whole country.*

75 *Economies of density should not be considered at the base station level but rather for a given geographical area. Indeed, when the traffic in a given area can no longer be supported by a single base station, the number of base stations in this area should be increased.*

76 *For example, during the construction of a 3G network, a 2G mobile operator can use part of its existing network, and approximately half the cost of facilities.*

77 *The controllers are the access points for all core network services. Station controllers allocate resources and control radio links. BSC (Base Station Controller) and RNC (Radio Network Controller) are the terminologies used for a base station controller in 2G (GSM) and 3G (UMTS) networks, respectively.*

- ***Core Network***

The cost of the core network is less important than the cost of the wireless local loop. It represents about 30% of total CAPEX, compared to about 70% for the wireless local loop. Excluding radio communications, the core network houses two types of network elements.

- Fiber optic lines

For this specific part of the network, we did not evaluate the degree of replicability, but rather assessed the essential aspect of the fiber optic lines for the deployment of a national mobile network. Our conclusion is that fiber optic lines are not essential to deploy a national mobile network. As such, fiber optic lines should not be considered in the ladder of investment and the decision to duplicate should be left to the operators.

While economies of scale can be realized in the deployment of fiber optic lines, they cannot be deemed as an important barrier to entry. First, the required traffic level to congest fiber optic lines is very high. In addition, in the case of dark fiber rental, since traffic management is not possible, the risk of anti-competitive behaviour from the infrastructure owner is low. Finally, the liberalization of the telecommunications sector has led to a significant number of replicated interregional lines. Competition in this segment is sufficiently strong to enable mobile operators to benefit from competitive wholesale prices. An entrant can, therefore, consider deploying a mobile infrastructure without building its own wireline network.

- Switching and subscriber control units (MSC, TMSC, HLR)

As they have limited capacity and since operating costs increase with traffic, switching and subscriber control units are not subject to economies of scale. Further, they are not specific to the mobile technology used (they can be used in both GSM and UMTS networks). Risks associated with this investment are, therefore, limited.

These elements of the core network are highly strategic, since this is where operations such as customer data collection are carried out (for example, HLRs provide traffic information per customer) and where call routing is controlled. With this type of strategic information, operators can differentiate their offers based on network usage and improve service quality (for example, with Fixed-Mobile Convergence, by routing calls from mobiles to fixed phones).

- ***Retailing***

Like the fixed broadband market, barriers to entry in this segment are low. Costs are for the most part variable, and neither investments nor expertise are specific to the telecommunications market.

Different entry thresholds exist across a mobile network. A number of tasks, for example, can be delegated to an existing mobile network operator (MNO): customer management and billing, sales, marketing or distribution. As such, market entry strategy may entail using a recognized brand in other markets, without making specific investments (e.g., the *M6 mobile* MVNO in France, from one the main private TV channels).

To build on our mobile market ladder in view of the fixed broadband market, we could have introduced a similar segment to the wireline local loop (the "last mile" in the fixed network). Based on this logic, the final phase could entail the purchase of a license to operate a radio frequency.

Our aim, however, was to build a ladder based on the replicability concept. As such, it would not have been relevant to include radio frequency as a ladder rung. Indeed, while a spectrum license is attributed by the regulator, replicability assumes that the investor has chosen between make or buy. We have consequently excluded the spectrum license from our mobile ladder.

Note, however, that the spectrum barrier influences the role of the regulator in the development of facility-based competition across the mobile market. If operators develop their infrastructures gradually before applying for a spectrum license, regulation should aim at encouraging entrants to climb the ladder. If, however, the entrant acquires a spectrum license prior to any investment, regulation is not incentive-focused since attribution is based on disclosure requirements for investment and territory coverage.

THE RELEVANCY OF THE LADDER OF INVESTMENT IN THE MOBILE MARKET

As stated by Cave (2004)[78], the objective of the ladder of investment is to enable entrants to gradually acquire customers and revenues. According to Martin Cave, the need for a gradual entry is linked to two problems faced by new entrants: the acquisition of a financial windfall and the creation of a customer base. It is interesting to analyze if these two elements, which are

78 *Cave (2004) argues that "[the entrants] may have to acquire capital assets progressively, as they acquire customers and revenues…"*

key in the concept of the ladder of investment for the fixed broadband market, retain their relevance in the case of the mobile market.

- ***The gradual acquisition of a financial windfall***

As investments necessary to deploy a mobile network (excluding the cost of the license) are lower than those needed to deploy a fixed broadband network, generating revenues to finance future investments seems less critical in the mobile market than in the fixed broadband market. Moreover, wireless local loop investments are the most costly. To develop this network segment, the operator must have a license, meaning that it has filed an application ensuring that it has sufficient financial resources.

- ***The gradual acquisition of a customer base***

The rapid acquisition of a customer base is fundamental in the mobile market, perhaps more than in the fixed broadband market, due to the importance of territorial coverage and club effects.

Mobile consumers deem territorial coverage as key. Aside from its license obligations, a new entrant is therefore obliged to provide an almost complete geographical coverage very rapidly to compete on level playing field with incumbent operators, which already offer a complete coverage. This is not the case in the fixed broadband market; the utility of a fixed broadband user does not depend on the geographical coverage of its broadband provider. Hence, for the mobile market, the concept of the ladder should be expanded to account for the geographical coverage of the operators.

In addition, in the context of club-based tariff plans,[79] acquiring market share is difficult for new entrants. Indeed, these operators have initially a small customer base and consequently do not enjoy the same network effects as incumbents. New entrants cannot, therefore, compete with the club-based tariff offers of incumbents.

79 *We refer to "club offers" as mobile contracts which offers lower prices for on-net calls than for off-net calls.*

REGULATORY TOOLS

➢ *Regulating access charges*

To motivate entrants to climb the ladder of investment, Cave suggests two forms of regulatory intervention based on the access charge of the existing infrastructure: a rising access price or the withdrawal of mandatory access at a given date (often referred to as the "sunset clause"). For Cave, the tradeoff between these two regulatory options depends on the degree of replicability of the given segment: "Where replicability is relatively certain however, mandated withdrawal of access may be a better approach" (Cave, 2006).

In a one-way access setting, Bourreau & Doğan (2006) show that these two regulatory tools are not equivalent. Sunset clause regulation implicitly assumes that the incumbent would charge the entrants too high a price when mandatory unbundling requirements are removed. If the incumbent historical operator prefers service-based competition to facility-based competition, it will however offer a relatively low access price as soon as access is no longer regulated to delay the advent of facility-based competition. In this case, the "sunset clause" would be ineffective.

Aside from this argument, the effectiveness of the "sunset clause" to force entrants to climb the ladder seems less certain in the context of the mobile market. Indeed, given that there are generally three or four mobile network operators, and if there is no longer mandatory access to the MNOs, competition between MNOs across the wholesale market might lead to competitive wholesale offers to entrants. Bourreau & al. (2007), however, show that competition between facility-based firms can lead to non-competitive prices in wholesale markets. If this is the case, lifting regulation of the wholesale market could provide entrants with incentives to climb the ladder.

Generally speaking, the market for access to the MNO networks is not regulated; in particular wholesale prices. What we observe could then be interpreted as the situation that would prevail after a sunset clause. MVNOs currently have a narrow path on which to compete with MNOs and depend on the whims of incumbents. Entrants are often restricted to niche markets, where dominant operators confine them by controlling wholesale prices and maintaining uncertainty through non-fulfillment contracts[80]. This suggests

80 *The contractual relationship between MNOs and MVNOs are unbalanced because of the lack of supervision of market 15 in the European Union (Call access and origination in public mobile telephone networks). Although the duration of the contract is generally short, an MVNO no longer benefits from competition once it has signed a contract with a MNO host. This is due to the high*

that an operator which has not taken the financial resources to develop its network could be locked in a niche, once mandatory access has ceased.

- ***Regulation of interconnection prices***

Depending on the rung of the ladder of investment, from which an entrant enters, it can or cannot receive interconnection revenues[81]. If the entrant receives interconnection revenues, with asymmetric interconnection charges[82], it receives an access benefit. To the extent that the financial windfall argument is relevant, this can facilitate the ascent of the ladder.

Interconnection prices for access to the entrant's network above marginal cost also lead to an increase of the average marginal cost of other operators and, hence, of the average marginal cost of the industry. As a result, the total quantity offered by the industry would decrease[83]. From a static point of view, this distortion would be socially costly.

This static inefficiency may be even greater if operators adopt a niche strategy, targeting a very specific clientele and therefore a small market share. In this case, the access margin could be viewed as a subsidy from dominant operators to virtual operators.

- ***Sequential award of new licenses***

In the mobile market, grants of new mobile licenses could provide an additional tool of public intervention. For instance, new licenses could be offered exclusively to new entrants or to late entrants (for instance, pure 3G players). Such a rule would provide new players with a first-mover advantage over incumbent operators for the development of new technologies or services and the acquisition of a customer base.

---

cost of switching suppliers (due to the change of SIM cards and terminals). This constraint is all the more important as entrants occupy a low position in the ladder of investment.

81 *For example, in the fixed broadband market, unbundled-based operators receive interconnection revenues, which is not true of operators using bitstream access wholesale offers. In the mobile market, only full MVNOs receive interconnection revenues.*

82 *Higher interconnection charges for access to the entrant's network than to the incumbents' networks.*

83 *In particular, the magnitude of the reduction of demand would be high with an elastic demand and if the asymmetry in market shares between the entrants and the incumbents were low.*

HOW THE MOBILE AND FIXED BROADBAND MARKETS DIFFER AND HOW THE DIFFERENCES IMPACT THE LADDER OF INVESTMENT

➢ *Spectrum: a scarce resource*

While fixed broadband operators use different types of access technologies (cable, power line…), the unbundling of the incumbent's copper local loop has developed service-based competition in most countries. In this context, the theory of the ladder of investment highlights that regulatory authorities ambition the development of competition between different access infrastructures. In the mobile market, the situation is different. Mobile network operators use a frequency band to deliver services. Since these frequencies are run by the States and represent a scarce resource, the number of competing infrastructures is technologically limited. Public authorities must consequently select which operators will receive a license and be allowed to compete in the market. Licenses generally entail obligations in terms of coverage and quality of service. To the extent that the path of development of their network is imposed in the schedule of conditions, the question of the optimal incentive mechanism to stimulate infrastructure development is obsolete in the case of a new licensed mobile network operator.

➢ *Duplication and social desirability*

As said above, facility-based competition is the long-term goal pursued by most regulatory authorities. Network duplication, however, has a social cost. Moreover, infrastructure-based competition in the mobile market raises public health issues and landscape preservation.

Before defining a ladder of investment, Cave (2006) insisted on the need to establish which network elements are clearly not replicable: "*If it is decided, for example, that the local loop is a natural monopoly, then any regulatory approach that creates infrastructure-building incentives in this element is not appropriate.*"

For mobiles, this replicability analysis should not only be based on cost structure, but must also incorporate a societal component (health, urban development…). Thus, even in a hypothetical case where frequencies are no longer a scarce resource, and where the number of MNOs would be unconstrained, encouraging too much duplication of the wireless local loop would probably not be optimal and not just for financial reasons. Infrastructure sharing would probably be an appropriate alternative to duplication for certain network segments.

One possibility would be to grant collective licenses rather than individual licenses. For example, a group of virtual operators could be awarded a collective license, encouraging them to share investments in the wireless local loop and to deploy their own core network.

CONCLUSION

Our analysis suggests that the concept of the ladder of investment could be applied, to some extent, to the mobile market. Indeed, a mobile network can be divided into different network elements, which can be classified according to their degree of replicability – as did Cave (2006) for the fixed broadband market. Moreover, one of the main reasons to implement a ladder of investment in the fixed broadband market - in particular, to facilitate the gradual acquisition of a customer base - seems equally relevant in the mobile market.

However, the ladder of investment for the fixed broadband market constitutes a regulatory response to a different situation than in the mobile market. Unlike the fixed broadband market that was (and sometimes, still is) characterized by a monopoly on the local infrastructure, in the mobile market, facility-based competition has been intense since market inception. Therefore, establishing an incentive mechanism, based on the concept of the ladder of investment, doesn't appear essential. Rather, the key issue is to understand why facility-based competition between MNOs does not lead to a competitive wholesale market for MVNOs. Bourreau, Hombert, Pouyet and Schutz (2007) suggest that the strategic interactions between the wholesale market and the retail market could explain this situation, and propose different ways of public intervention.

There are also technological obstacles to establish a ladder of investment for the mobile market. Radio frequencies constitute a limited resource. The number of players which can climb the ladder entirely is therefore technologically limited. Thus, an operator investing in the construction of a mobile network up to a given ladder rung would face a high degree of uncertainty, because the final investment in the wireless local loop depends on available frequencies.

Finally, infrastructure duplication in mobile markets is not necessarily desirable, not only for standard cost considerations, but also for public health considerations or the preservation of landscape. The social desirability of a ladder of investment in the mobile market does not therefore appear as clear-cut as for the fixed broadband market.

If, however, we ignore these remarks and assume that the ladder of investment is relevant in the mobile market, it would be important to account for fixed-mobile convergence. Indeed, players in the fixed broadband market

have to integrate a mobile offer[84] to provide a diversified and competitive range of services. Operators having climbed the ladder of investment in the fixed broadband market benefit from economies of scope if they enter the mobile market. Indeed, the infrastructure that these operators own in the core network – e.g., transmission links – could be used to develop a mobile infrastructure, leading to substantial investment savings. These facilities could also be rented to other MNOs or MVNOs, and hence generate wholesale revenues.

In addition, owning an SIP network allows these operators to provide users with a mobile access to their network. Some fixed broadband service providers could also use WIFI coverage provided by their ADSL box to benefit from a wireless communication network, such as mobile VoIP offers. This opportunity has two benefits: it would allow the convergent operator to depend less on host MNOs in areas where the operator has a strong capillarity of ADSL users and to provide more attractive VoIP offers.

REFERENCES

Bourreau M., Doğan P. (2006), "Build or Buy Strategies in the Local Loop", American Economic Review, Papers and Proceedings, Vol. 96, pp. 72-76.

Bourreau M., Hombert J., Pouyet J., Schutz N. (2007), "The Competitiveness of Intermediate Goods Markets with Vertically Integrated Firms", Mimeo.

Cave M. (2006), "Encouraging Infrastructure Competition via the Ladder of Investment", Telecommunications Policy, Vol. 30, pp. 223-237.

Cave M. (2004), "Making the Ladder of Investment Operational", Paper presented to the European Commission, November.

European Commission (2002), "Comparative Assessment of the Licensing Regimes for 3G Mobile Communications in the European Union and their Impact on the mobile Communications Sector", June 25th.

84 Conversely, mobile operators will likely try to enter the fixed broadband market.

Comparison of Structure, Conduct and Performance: U.S. versus Europe's Wireless Markets

JOSEPH P. FUHR JR.
Widener University and the American Consumer Institute

STEPHEN POCIASK
The American Consumer Institute

Abstract: *There is a perception, by some, that the U.S. wireless market, compared to its European counterparts, is failing and in need of regulatory remedies. This chapter tests this hypothesis by examining the evidence of market failure and finds that the U.S. wireless market has more choice, lower concentration, higher monthly usage and lower prices than in other parts of the world. In effect, increased facility-based competition appears to go hand-in-hand with increased consumer welfare. Therefore, the contention that the U.S. wireless market lags its international counterparts is not supported by economic data and no further regulation is needed to address this issue.*

INTRODUCTION

A frequent claim is that the European wireless market has more choices, is healthier and has lower prices than the U.S. wireless market. For many years, statistics from the Organisation for Economic Co-operation and Development (OECD) have shown Europe to be far ahead of the U.S. in terms of cellular service penetration. Suggestions that the U.S. wireless market is slower to innovate and generally lags other parts of the world have contributed to a recent Federal Communications Commission (FCC) decision to place additional rules upon bidders in the 2008 wireless spectrum auction. In testimony before the U.S. House of Representatives, senior counsel for Consumers Union testified that "in Europe and Asia, wireless consumers have better choices" (Murray 2007). Referring to the U.S. market, this testimony states, "instead of innovating, the wireless industry has become a cozy cartel of a few dominant providers with limited device offerings" (Murray 2007). All of these concerns have fueled a debate over the need to correct the U.S. wireless market, including proposals to add regulations, mimicking a European-style model.

If there is a problem, a healthy public policy debate is needed to address how to fix it. However, before coming to conclusions that regulatory remedies should be considered, a review of the empirical evidence is needed. The purpose of this chapter is to examine whether the U.S. market actually lags Europe and, if so, the extent of the gap.

EVALUATING MARKET FAILURE

Economic theory provides a rationale for regulation but only if markets fail and the remedy addressing this market failure (e.g., government regulation) leads to net social benefits. Therefore, in this study, the contention that the U.S. lags Europe will be evaluated by comparing the wireless industry's structure, conduct and performance in search of actual market failure.

The contention that the U.S. lags Europe is supported by the following hypothesis – the U.S. market is more concentrated than the European market, leading to higher consumer prices and lower penetration rates in the U.S. Based on this purported market failure, the end result is lower consumer welfare in the U.S. than in Europe. This is a testable hypothesis.

MARKET STRUCTURE

Data from the OECD and FCC suggest that the U.S. wireless market is not as concentrated as markets in Europe, or elsewhere for that matter. Data from the OECD show that the U.S. wireless market has a higher number of wireless competitors than in Europe or any other county. Therefore, according to publicly available information, the conclusion that Europe has more competitors and more choice is incorrect. This point is also supported by a Merrill Lynch (2007) report, as well as several consulting reports (Schwartz and Mini 2007, Ford & al. 2007, and Lowenstein 2007). Further, a CTIA (2007) report shows that the U.S. offers 700 different wireless handsets compared to about 190 for the U.K.

Looking at the number of wireless operators "permitted" to provide wireless broadband services, the U.S. leads all European counties. Moreover, the "U.S. mobile operators have the flexibility to upgrade their networks to 3G on their existing 2G networks (cellular, PCS, SRM) spectrum," (OECD 2007) while European GSM operators must rebuild their networks. This means that the OECD data understate the number of permitted U.S. operators, and explains how the U.S. is rolling out wireless broadband services so quickly and sometimes offering services that provide twice the speed of European wireless broadband services (Mossberg 2006).

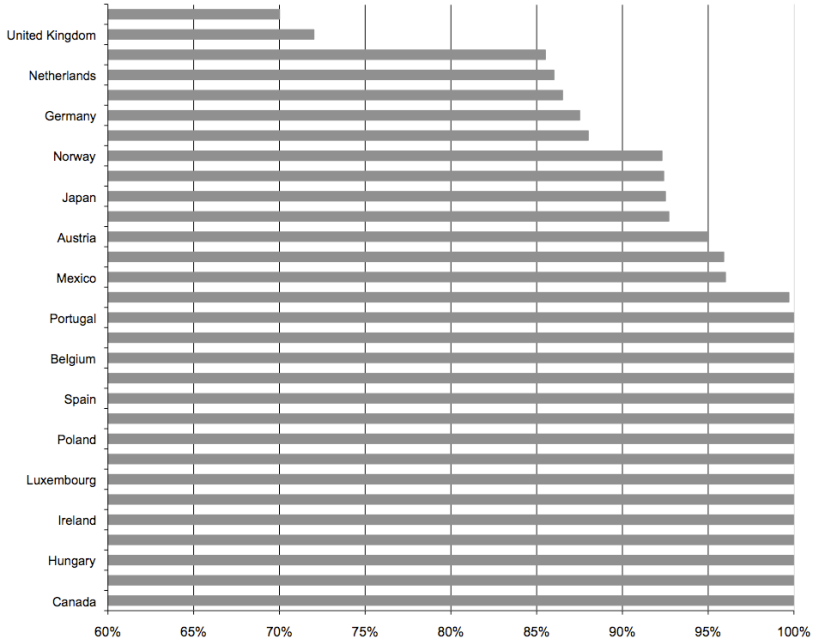
Table 4. Number of Wireless Operators (2005)

	MNO	Licences
Australia	4	4
Austria	4	4
Belgium	3	3
Canada	17	2
Czech Republic	3	3
Denmark	4	4
Finland	15	3
France	25	3
Germany	4	4
Greece	4	3
Hungary	3	3
Iceland	3	0
Ireland	4	3
Italy	3	4
Japan	17	12
Korea	3	3
Luxembourg	3	3
Mexico	18	1
Netherlands	4	4
Norway	3	3
Poland	3	4
Portugal	3	3
Slovakia	2	2
Spain	3	4
Sweden	4	4
Switzerland	5	4
Turkey	3	0
United Kingdom	5	5
United States	155	5+

Source: OECD Communications Outlook, OECD, 2007, p. 35, 2005 Data

Is it possible that the U.S. market has more wireless operators, but it is dominated by only a few? The OECD data give the market share for the top 3 wireless providers and show that the contention that the wireless market is a “cozy cartel” is less a U.S. problem than it is a European one.

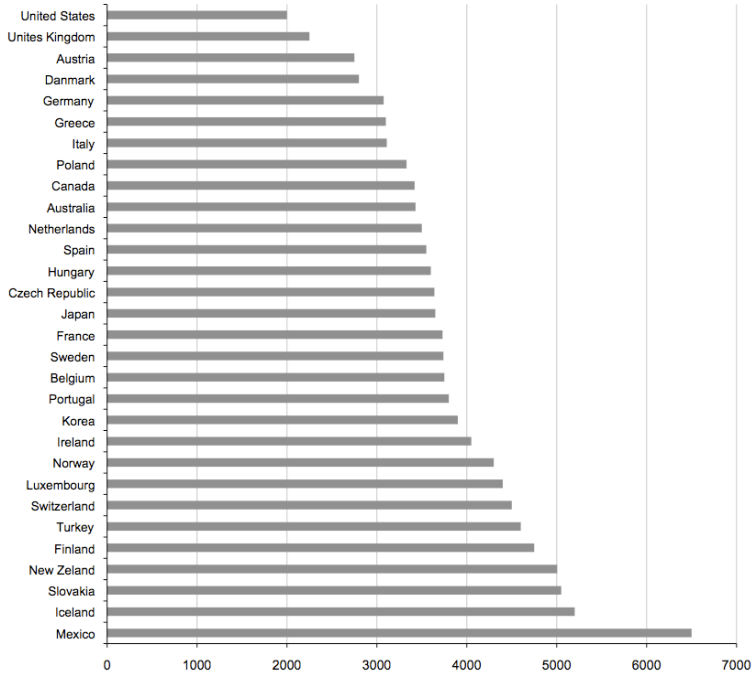
Figure 6. Market Share - Top 3 Firms



Source: OECD, ACI Calculations, 2005

A final way to look at industry concentration is to use the measure widely accepted by the U.S. Department of Justice, U.S. Federal Trade Commission and many world courts – the Herfindahl-Hirschman Index. We find that the U.S. has the lowest concentration index when compared to all of the OECD listed companies. In summary, the U.S. wireless market has more operators and is less concentrated than any other international market for which data are available.

Figure 7. HHI concentration indexes

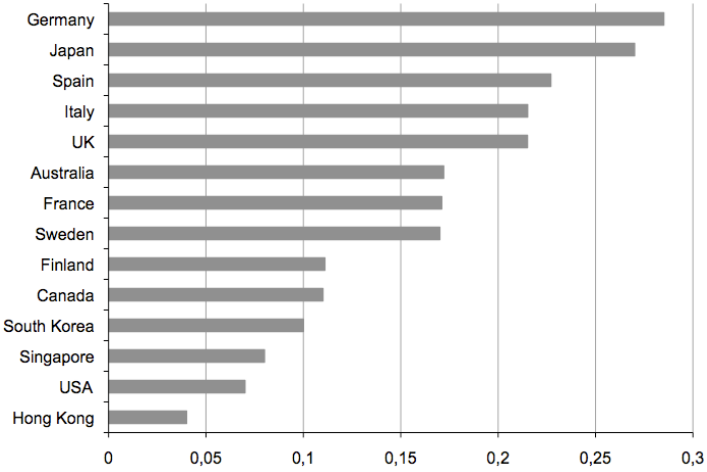


Source: OECD, ACI Calculations, 2005

CONDUCT AND PERFORMANCE

If the U.S. market is not as concentrated as its European counterparts, could it still operate as a "cozy cartel," effectively driving up consumer prices and restricting supply? This is also a testable hypothesis. FCC data (next figure) show that the U.S. has the lowest rates (measured by revenue per minute) of any major developed economy with the exception of Hong Kong. The FCC concludes "mobile calls continue to be significantly less expensive on a per minute basis in the United States than in Western Europe or Japan" (FCC 2006). Therefore, not only is the U.S. less concentrated, but it offers more competitive prices than other countries.

Figure 8. Average Revenue per Minute ($)

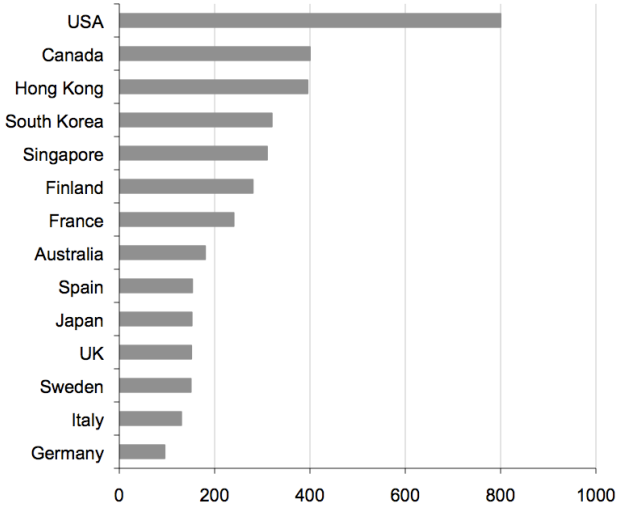


Source: FCC

It is important to note that Europe uses a "calling party pays" system that bills customers for originating minutes, whereas, in the U.S. system, customers are billed for both originating and terminating minutes. This means that the U.S. system counts more minutes, which would affect the relative comparison of average monthly usage and price. Schwartz and Mini (2007) conducted an independent analysis correcting for differences in payment system and still found U.S. prices to be lower than all of the major European countries. They also make a correction for the prevalence of dormant (not active) phones in the European statistics. Their analysis also shows that U.S. prices have decreased faster as a percent or in absolute terms, compared to all major European countries (Schwartz and Mini 2007).

Another adverse effect of market power is restriction of supply which leads to consumer welfare losses. However, data from the FCC show that the U.S. market has the highest usage. In 2005, U.S. wireless consumers talked on average 800 minutes per month, while consumers in some European nations average less than 200 minutes. In other words, the more concentrated European wireless markets charge consumers substantially more per minute which, in turn, leads to lower usage by consumers.

Figure 9. Wireless Minutes of Use per Month

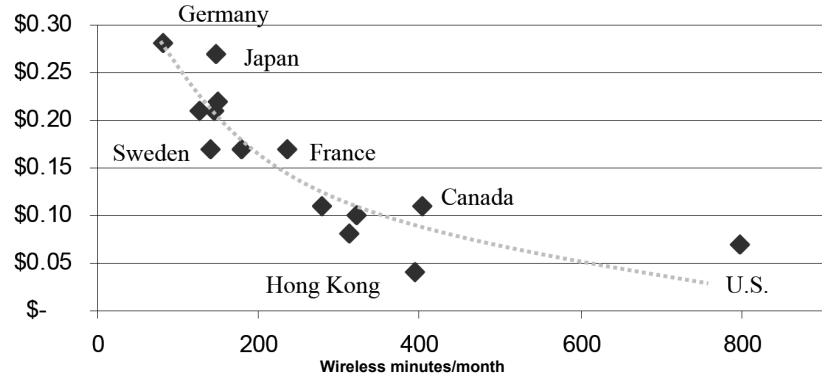


Source: FCC

These conclusions hold up even after accounting for differences in measurement between Europe and the U.S. Specifically, Schwartz & Mini (2007) show that the average U.S. consumer uses about three times more minutes per month than the average European consumer.

Another comparative observation is that the relationship between usage and price appears to be negatively correlated. As depicted in the chart below, as wireless prices increase, customer minutes decrease. U.S. consumers evidently benefit (as measured by consumer welfare) more from wireless services than their European counterparts. From these data, we conclude that market failure appears to be a European problem, not a U.S. problem.

Figure 10. More Minutes at Lower Prices



Source: OECD

NON-PRICE EFFECTS

This chapter has discussed how increased facility-based competition appears to go hand-in-hand with lower consumer prices and increased consumer welfare. However, while it is more difficult to measure, some anecdotal evidence suggests that increased competition also leads to non-price effects.

In general, increased facility-based competition in the U.S. market has made wireless providers very responsive to changes in consumer preferences. Considering that it is consumers that pay the bills and, when given market choices, anecdotal evidence suggests that wireless providers are indeed listening. Wireless consumers are now expressing more diverse needs and there are indications that competitors are responding by increasing the differentiation of services in order to meet those needs. Pricing solutions that attracted mainstream consumers years ago and standardized “out-of-the-box” solutions are slowly giving way to demands for specialized solutions, flexibility and service differentiation.

Facility-based competition has led to continual improvement in service quality. For example, the Federal Communications Commission reports that wireless complaints fell by 33% in 2006. Faster wireless high-speed Internet services are now available to hundreds of millions of consumers and consumers are demanding more, better and faster services, as evidenced by the frenzy surrounding the iPhone.

In addition, competitors are now offering prorated early termination fees and are unlocking phones, which enables consumers to switch more easily between competitors. Wireless competitors are now opening their network to

allow access to any web site, as well as allowing the use any devise or applications. There are also many nuances that may attract some consumers to one service provider over others. For example, some providers offer flexible rollover minutes and most providers offer 30-day no-question-asked return policies. As other competitors enter the market, service innovation and differentiation are likely to continue – all to the benefit of consumers and all without onerous regulations. In short, competition has given consumers the power and the market is listening.

ONE CAVEAT: POTENTIAL PROBLEM WITH OECD PUBLISHED DATA

This chapter has shown that the U.S. market has lower market concentration, more competitive prices and greater acceptance among consumers. However, these findings are counter to published international data on market penetration. The problem with these international statistics is that they do not count the same things. In the U.S., wireless subscriptions primarily reflect the number of handsets in operation. However, in Europe, the statistics reflect the number of Subscriber Identity Module (SIM) cards. Since European roaming rates are so high between countries, it is sometimes economical for consumers to have more than one SIM card per handset, to take advantage of lower in-country rates. As a result, one wireless customer may have several SIM cards and be counted multiple times in the OECD statistics. In addition, pre-paid wireless plans are more prevalent in Europe than in the U.S., and these customers tend to use wireless services much less than post-paid customers. Correction of these differences will decrease the reported European penetration rate relative to the U.S.

CONCLUSION

Calls for a regulatory remedy are not based on sound empirical evidence. The data from the OECD and FCC show that European wireless markets have higher concentration, higher prices and lower usage. From this analysis, the U.S. wireless market gives consumers more choice, offers more competitive prices and encourages more consumption. Compared to Europe, reported problems of high market concentration, high consumer prices, low usage and decreasing consumer welfare do not appear to be a U.S. problem. In summary, there is no evidence of market failure or that the U.S. wireless market somehow lags behind the European wireless market. In fact, if anything, basic comparisons of consumer welfare between these markets demonstrate the opposite conclusion, perhaps suggesting that

increased facility-based competition may be one key to the market's success.

REFERENCES

CTIA (2007), "CTIA Ex Parte Communications with the FCC", WC Docket No. 07-52, July 24th.

FCC (2006), "Eleventh Wireless Report, FCC", released September 29th.

Ford G., Koutsky T., Spiwak L. (2007), "Wireless Net Neutrality: From Carterfone to Cable Boxes", Phoenix Center for Public Policy, Bulletin No. 17, April.

Lowenstein M. (2007), "Comparisons between U.S. and European Markets for Wireless Services and Devises: Myth vs. Reality," July, submitted in Ex Parte communications with the FCC, WT Docket No. 06-150, July 23th.

Merrill Lynch (2007), "Global Wireless Matrix 1Q07: Latin Fever", Merrill Lynch, June 15th.

Mossberg W. (2006), "Cingular Joins Rivals with Fast, Reliable Wireless Broadband", Wall Street Journal, January 19th.

Murray C. (2007), "Wireless Innovation and Consumer Protection, Testimony before the U.S. House of Representatives on Telecommunications and the Internet", Committee on Energy and Commerce, July 11th.

OECD (2007), "Communications Outlook".

Schwartz M., Mini F. (2007), "Hanging up on Carterfone: The Economic Case Against Access Regulation in Mobile Wireless", May 2nd.

Optimal Mobile Telephony Market Structure in Europe: "Two are too Few and Four are too Many"?

LAURENT BENZONI
University of Panthéon-Assas

PATRICE GEOFFRON
University of Paris-Dauphine

***Abstract**: A comparative analysis of European mobile market structures (in terms of number of operators) and performances (price, volume, penetration) is presented, in order to reveal an "optimal structure". It has been shown that markets with three players are those that have the most success in relation to these criteria. Without concluding that this is a "golden number", these results suggest to authorities that there is not necesseraly an economic rationale to add more operators (MNOs as well as MVNOs) that will have to face the persistant first movers advantages of the incumbents. It also indicates that granting an additional license doest not lead, in any circumstance, to an increase in the consumers' surplus.*

INTRODUCTION

The "structural" approach in industrial economics stipulates (roughtly) that the intensity of the competition is inversely proportional to the concentration of the offer. However, in the presence of scale economies, the unit costs are decreasing with the level of production (number of users and volume of traffic in the case of mobile phones). In a market where the demand is fixed, an increase in the number of firms limits their individual size and, thus, thwarts economies of scale. Thus, the dispersion of the offer (that guarantees the competitive intensity) therefore hinders productive effectiveness.

In the various European mobile markets, the number of infrastructure operators varies from two (as in Norway, for example) to five (as in the United Kingdom). Faced with this diversity in national market structures, it would be helpful to empirically determine whether a "natural structure" to the offer is observed over the long term. In other words: can European experience lend itself to demonstrating an optimal structure of the mobile telephony market that balances, as much as possible, the contradiction between large-scale players' needs (for economies of scale), and the necessity for numerous players in order to preserve competitive pressure?

Our paper is designed to highlight this question, within the context of European mobile telephony. The first section will address the concept of the "optimal market structure", in a resolutely simple and didactical manner. The second section analyses the data relating to Europe's mobile telephony markets (in terms of price, volume and rate of penetration), enabling the optimal structure to be assessed, and, most specifically, providing evidence for the debate concerning competition in infrastructures within the mobile telephony markets.

On the basis of at least ten years' experience in mobile markets, what can be said concerning the number of operators for whom granting an additional license no longer benefits the consumer?

"BACK TO BASICS": THEORETICAL REMINDERS OF THE OPTIMAL MARKET STRUCTURE

Within the context of a market open to competition, the number of firms is theoretically not limited (except by the profitability of entry), enabling various market structures, competitive intensities and levels of equilibrium prices to be envisioned. Theoretically, once again, if the competition is not biased, the increasing the number of firms on the market will cause the prices to lower.

The standard equation therefore stipulates that:

$$L = \frac{HHI}{e}$$

Where:

- ***L*** = The Lerner Index of market power[85].
- ***HHI*** = Concentration Index[86].

85 *L= (P-Mc)/P where P = price and Mc = Marginal cost. Thus, the Lerner Index is greater than 1 and increases as the price exceeds the marginal cost (denominator in the equation). The market power is therefore measured as a firm's or group of firms' capacity to raise the price above marginal cost. In the standard model, equality between the price and marginal cost is the pure and perfect balance of competition.*

86 *The HHI (Hirshman-Herfindahl Index) is a concentration index. It is calculated by adding the squares of market shares for all firms in the sector. For example, in a market with five firms who each have a market share of 20%, the HHI is: 400 + 400 + 400 + 400 + 400 = 2000. The higher the HHI is, the higher the concentration, with a maximum value of 10,000 reached for a monopolized market. The HHI value rises when: 1/ The number of firms present is low; 2/ The dispersion of the sizes of firms in competition is high.*

- ***e*** = Price demand elasticity.

As a result, the higher the concentration, the greater the market power is, and the greater the risk that there will be a larger gap between the prices and the level of competition, which guarantees collective welfare. However, the maximum number of viable players over the long term on a market may be limited by high fixed costs. If the number of providers present on the market increases, the average market share will mechanically decrease, as well as the scale economies.

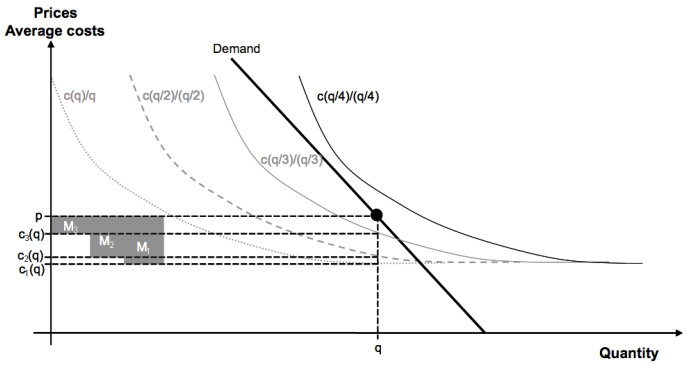
Let us consider a standard operation of average cost:

$$Cost_{Unit} = \frac{Cost_{Fixed}}{Quantity} + Cost_{Variable}$$

Assuming that all operators use the same technology over the long term (efficient operators), let's outline the cost functions for diverse market structures with 1, 2, 3 or 4 players, assuming that each has an equal market share.

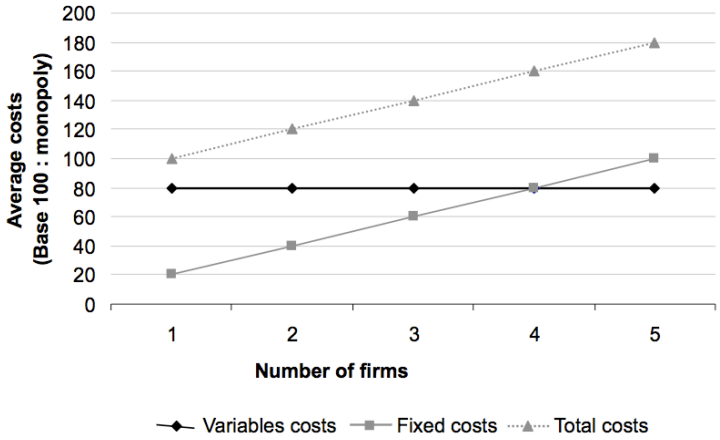
The following figure shows (classically) that when there are few providers present on the market, the profit is very significant. However, exceeding a maximum number of players, this positive profit disappears and the long-term survival of an "n^{th}" operator cannot be assured (in the present case, n=4). On the short-term, the providers in place might prefer making room for this "n^{th}" provider, which in turn leads to an increase in prices, rather than risking a "price war" with an uncertain outcome which furthermore would not increase the margin level.

Figure 11. Evolution of Costs and Margins in Terms of Market Structure



To clarify even further, the next figure indicates the unit cost of a product, according to the number of providers present on a market, assuming that these providers have equal market share, and that the fixed costs in the monopoly case represent 20% of the total costs. It is noted that, in this case, the unit costs increase significantly according to the number of providers on the market.

Figure 12. Evolution of average costs in terms of the number of providers on the market



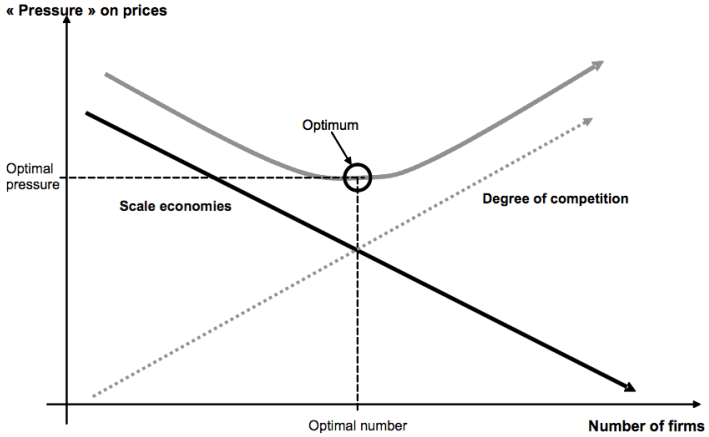
Inasmuch as the unit costs are more significant within the context of a market with many firms, these firms have the tendency to compensate for these costs through prices.

Thus, two contradictory forces act according to the number of players present:

- competitive pressure causes prices to decrease when the number of players increase...
- …and economies of scale when the number of players decreases.

Ideally, the number of players on the market that poses the best compromise between these two forces should be found.

Figure 13. Optimal market structure

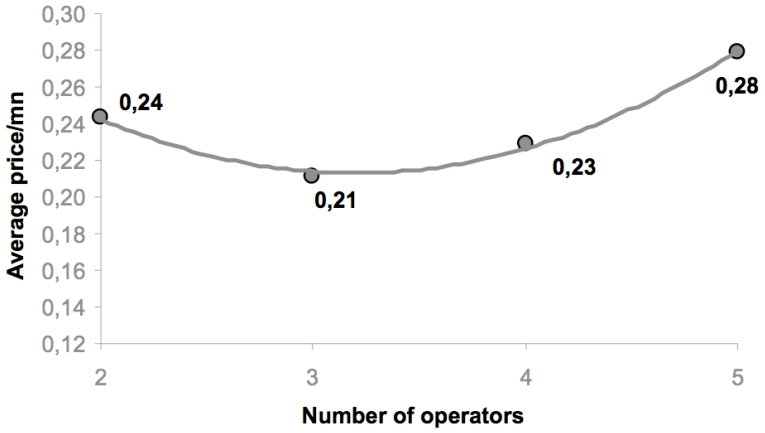


TRANSPOSITION TO THE MOBILE TELEPHONY MARKET IN EUROPE

Concerning mobile telephony, managing this pressure is the responsibility of the public authorities who assign frequencies (thus designing the market structure), and not the result of a natural entry/exit process. Modelling the simultaneous impact of the competitive pressure and economies of scale on the prices of mobile telephone is not trivial. However, it is possible to outline a comparative analysis that crosses the market structures (from 2 to 5 operators in the European case) and performance indicators.

The following figure shows the average price per minute in European countries according to the number of operators present on the national market. Thus, consumers in countries with a 3-operator structure profit from lower average prices – about 0.21€ per minute. Countries with a more concentrated 2-operator structure, or those on the opposite end of the spectrum with a less concentrated 4- or 5-operator structure have higher average prices, from 0.23 to 0.28 €.

Figure 14. Average price per minute, according to the number of operators on the national market (in 2005)[87]

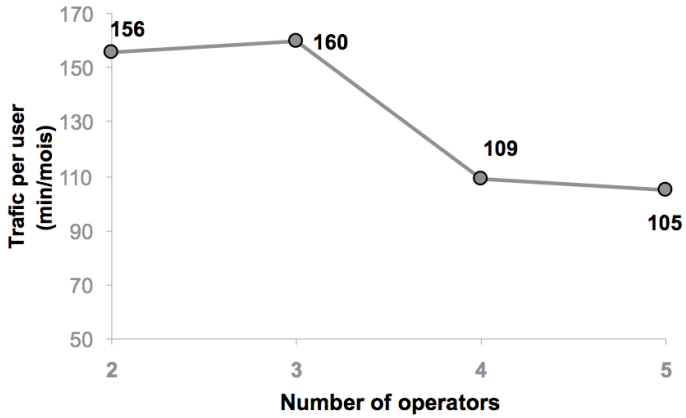


Source: OMSYC-Quantifica data, TERA analysis

It is noteworthy that, coherent with the price level, average consumption per user (traffic in minutes) is maximized with the market structure of 3 operators. In fact, the average consumption per user in countries surveyed with 3 operators is 160 minutes per month, a number that falls to 109 minutes per month with 4 operators, representing a decrease of more than 31%. With 5 operators, average consumption falls to 105 minutes per month. It is also lower on a market with 2 operators.

87 *The sample is composed of markets of the EU 15, excluding Luxemburg. Prices are calculated in PPP.*

Figure 15. Average traffic per user, according to number of operators on the national market (in 2005)



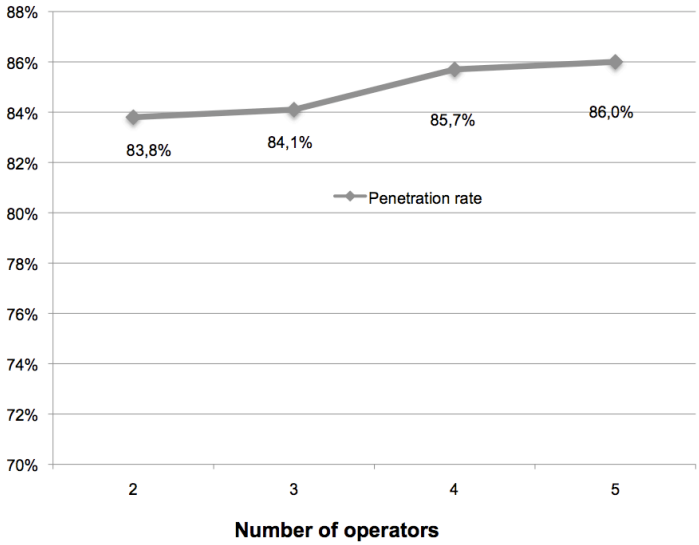
Source: OMSYC data, TERA analysis

The comparative analysis of 14 European countries thus reveals that the markets with a three-operator structure present both lower average prices per minute, and higher average consumption per user.

Some may object, stating that this relationship between structures and performance could be explained differently. In fact, in countries with a high number of operators (low concentration), the competitive dynamic could foster more widespread use of mobile phones throughout the population. Therefore, the greater the distribution, the more the "small" consumers must be reached. As these small consumers do not benefit from the best unit rates, the average price per minute of communication sold on the market may rise, while average traffic per user may decrease.

To verify whether or not this phenomenon exists in Europe, the number of operators and penetration rate of mobile telephony within the population must be compared. From this, it is deduced (refer to next figure) that the increase in number of operators finally has almost no impact on the penetration rate; increasing from 3 to 4 operators only provides a 1.6% increase in penetration rate, while increasing from 4 to 5 operators only provides a 0.3% increase.

Figure 16. Average penetration rate of mobile telephony, according to the number of operators on the national market (in 2006)[88]



Source: OMSYC-Quantifica data, TERA analysis

Therefore, the higher rate levels and lower consumption volumes in countries where there are 4 or 5 operators can not justify the differences in penetration of mobile telephony within the population.

COMMENTS AND CONCLUSIONS

It is important, obviously, to produce very prudent comments on the basis of these results. First of all, because the sample is reduced, and secondly because factors other than prices or quantities shall be taken into account to judge performances (innovations in offered services, most specifically).

We will therefore formulate a very careful conclusion:

- The data gathered suggest that the equilibrium point between pressures ("economies of scale" vs. "competitive intensity") would be 3 operators.

88 *Based on the penetration rate of individuals equipped within the population, and not on the SIM card penetration rate, which neglects the phenomena of double or triple equipment.*

- On the contrary, and proof of our caution, nothing indicates that the 4th and 5th operators produce a perceptible positive effect for the user.

To briefly refer to the famous Selten's article (1973)[89], this present report suggests that, on the European mobile markets, "*2 operators are too few, and 4 are too many*".

This should incite reflection relating to the European Commission's policy that attempted to promote market structures with four or five players when third generation licenses were granted, and to any regulatory authority that would like to believe that, in a sector of infrastructures such as mobile networks, no matter what the circumstances, granting additional licenses is a gauge of the increased collective long-term welfare. Benzoni & Geoffron (2007) have produced a detailed analysis of the "first movers advantages" (i.e. of the first license holders) that explain this phenomenon and highlight the low market power of the 3th, 4th, 5th operators.

The European history of 2nd generation mobile markets therefore bears the mark of a competition infrastructure biased by awarding of licenses over time, which in turn "cursed" the most recent entries. Though history has been written, the future is open, under the following conditions :

- Authorities shall not consider that service-based competition (through MVNOs) suffices to correct the initial bias of infrastructure competition. It is likely that, no more so than the most recent operators who've entered the market, the MVNOs will not be able to compete with incumbent operators. It is therefore not guaranteed that MVNOs is an efficient to "tool" to manage pressure between "economies of scale" and "competition".
- In addition, it is important not to reproduce the errors experienced in the 2nd generation (GSM), for the 3rd (UMTS) and 4th generations. Even if it may already be too late for the UMTS.

89 *Selten, R., 1973, "A Simple Model of Imperfect Competition where 4 are Few and 6 are Many", International Journal of Game Theory, 2, 141 – 201. Recent economic experiments place the same bar below Selten's. Case in point: Huck S. & alii (2004), "Two Are Few and Four Are Many: Number Effects in Experimental Oligopolies", Journal of Economic Behavior & Organization, 53, 435–446.*

REFERENCES

Benzoni L., Geoffron P. (eds.) (2007), "Competition and Regulation with Asymmetries in Mobile Markets", Quantifica, 2007.

Busse M. (2000), "Multimarket Contact and Price Coordination in the Cellular Telephony Industry", Journal of Economics and Management Strategy, 9, No. 3, Fall, 287-320.

Gebhardt G., Wambach A. (2008), "Actions to Implement the Efficient Market structure", International Journal of Industrial Organization, Volume 26, Issue 3, May 2008, Pages 846-859.

Heil O.P., Helsen K. (2001), "Toward an understanding of price wars: Their Nature and How They Erupt", International Journal of Research in Marketing, 18 2001 83–98.

Huck S. & alii (2004), "Two Are Few and Four Are Many: Number Effects in Experimental Oligopolies", Journal of Economic Behavior & Organization, 53, 435–446.

Klemperer P. (1989), "Price Wars Caused by Switching Costs", Review of Economic Studies, 56, 405-420.

Parker P.M, Röller, L.H. (1997), "Collusive Conduct in Duopolies: Multimarket Contact and Cross-Ownership in the Mobile Telephone Industry", Rand Journal of Economics 28, Summer, 304-322.

Selten R. (1973), "A Simple Model of Imperfect Competition, where four are few and six are many", International Journal of Game Theory, 2, 141–201.

Stigler G.J. (1964), "A Theory of Oligopoly", Journal of Political Economy, 72, 44-61.

The "Service versus Infrastructure" Debate in the Context of the European Mobile Markets

CUONG HUNG VUONG
University of Paris II (Panthéon-Assas)

Abstract: *This paper discusses the reasoning favouring infrastructure over service-based competition in the context of the European mobile telephony sector. Using principally basic economics and finance toolkits, it is stressed that, because of the characteristics of mobile industry (a fast changing industry with high initial irreversible investments) and the complex nature of competition in the European mobile industry, infrastructure-based competition presents more "guarantees" as regards innovation dynamics over the long term.*

INTRODUCTION

The controversial issue of service-based competition versus infrastructure-based competition in telecommunications has retained the attention of economists, lawyers as well as regulators. Service-based competition is essentially believed to bring lower retail prices, more innovation at the "edges", while infrastructure-based entails more investments, and hence more advanced technology and innovations at the "core", including higher cost reduction (Bergman 2004, Henten & Skouby 2006).

The same opposition of views re-appears in mobile services regarding competition between MVNOs (Mobile Virtual Network Operators) and MNOs (Mobile Network Operators). Some believe that there should be intervention in mobile markets to encourage entry of MVNOs to increase competition on the retail market and consumer choices, while others suggest intervention should be minimal.

Certainly, these conflicting interests of both parties may call for appropriate regulation to harmonise both types of competition, when possible. To address this issue, we argue that because of the characteristics of mobile industry (a fast changing industry with high initial irreversible investments) and the complex nature of competition in this kind of industry in Europe (both national and pan-European players, first movers advantages,…), infrastructure-based competition presents more "guarantees" as regards the innovation dynamics over the long term.

OVERALL ASSESSMENT OF INFRASTRUCTURE COMPETITION IN THE EUROPEAN MOBILE INDUSTRY

It is obvious that mobile telephony is a fast changing industry, starting with analogue technology, following into the 1990s with the GSM technology boom, perceived as a successful regulatory approach in Europe[90]. At the current time of writing, many European MNOs have already rolled out 3G networks with much higher data transmission speeds, more powerful platforms, creating greater room for innovation and providing greater benefits for mobile users. On the cost side, the mobile industry is characterised by low marginal costs but huge initial investments to set up networks (base station level for example). More importantly, most of these investments are irreversible (for example the cost of digging and installing the lines to roll out new networks).

This clearly leads to a number of striking implications:

- First, from a social point of view, irreversible investment can be seen as an instrument of business commitment.
- Second, scale economies exist in the mobile industry; i.e.: the more customers served leads to expected lower retail prices, although retail prices will never equal marginal costs.
- Finally, due to market uncertainty, the risks associated with investments are generally higher than they are in other industries where costs can be reversed.

In European mobile markets, infrastructure-based competition was introduced from the very beginning with around two competitors per country. Today, in each European market, there are between 3 and 5 infrastructure-based mobile competitors. Thus, the issue of "slow introduction" of infrastructure-based competition (like in broadband) has not been a real issue for mobile markets because many mobile operators have built their networks in less than 2 years. Moreover, it seems that, in choosing to promote such infrastructure competition there has been no or few inefficient duplications on investments (Benzoni and Geoffron 2007). Even if, at the lowest level (local network), the efficiency of Base Stations (BTS) duplication might be questioned, site sharing offers an opportunity to inhibit this potential problem.

90 *For a detailed history of Europe's mobile telephony, one can consult Dunnewijk & Hultén (2007).*

The positive results of infrastructure competition may be summarised in considering, as detailed in the next figures :

- The evolution on penetration rate up to 110% in 2007.
- The 50% decrease of prices between 2004 and 2007 for a medium consumer[91].

Figure 17. Evolution of mobile services prices in Europe

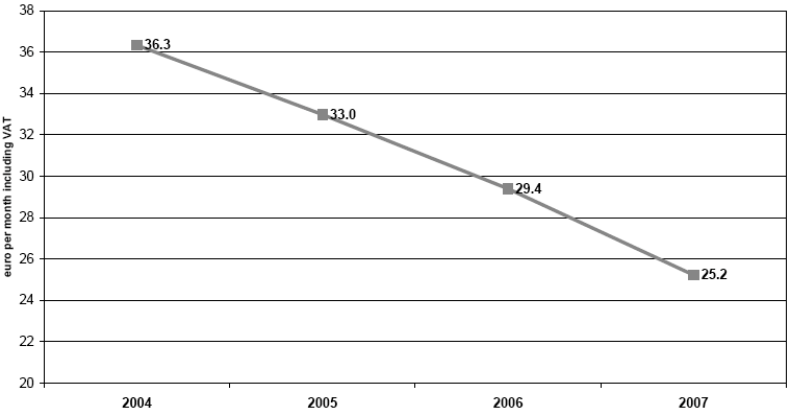
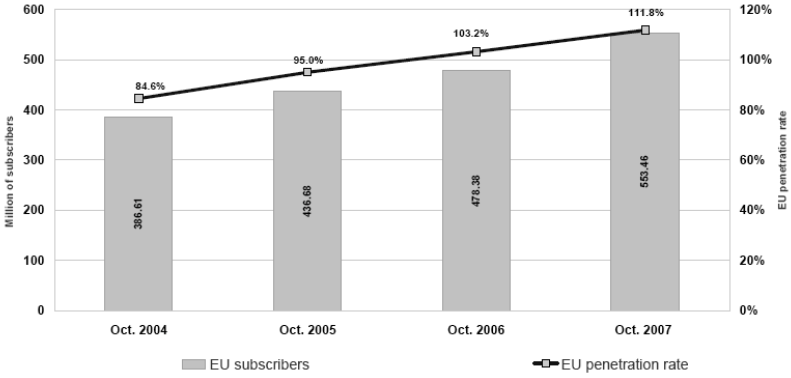


Figure 18. Evolution of mobile penetration rate in Europe



Source: European Commission

91 *European Commission, "European Electronic Communications Regulation and Markets 2008"*

Thus, such figures are consistent with analyses developed in the broadband where, Dispaso *et al* (2006) and Wallsten (2006) show that infrastructure-based competition ('inter-brand' competition) is more likely to increase the diffusion rates rather than service-based competition ('intra-brand' competition). And, we shall consider that these results have been obtained with a rather low regulation level as regulatory bodies just intervene in subjects such as interconnection, national roaming or number portability. In mobile markets, regulation in Europe is far less strict than in fixed markets as there was more than one GSM operator from the beginning.

Nevertheless, some limits of the infrastructure competition must be underlined.

A fallacy in competition analysis is to basically observe the number of players in an identical market while the nature of competition has its own characteristics. In the mobile industry, due to fast changing technology, MNOs are likely to engage in competition *"for"* the market, as well as *"in"* the market and benefit from "first mover advantages". This can be evident from the current European mobile market using GSM technology, where late entrants remain much smaller than the earlier entrants in terms of market share and profitability (Benzoni and Geoffron 2007).

The introduction of more service-based competition through MVNO, may weaken moreover the later entrants without limiting the incumbents' market power. For example, compared to the previous year, the number of mobile service providers in EU (mobile virtual network operators, enhanced service providers or simple resellers) has risen to 290 in 2006, an increase of 76. Nevertheless « *the actual decrease in the market share of the leading operators was relatively small between 2004 and 2006. In percentage terms, the decrease in the leading operator's market share in terms of subscribers between 2005 and 2006 was only 0.2% while the main competitors increased their market share by 0.1% between 2005 and 2006* » [European Commission 2007].

CONSIDERATIONS ON THE OPTIMAL REGULATION OF MNOS-MVNOS RELATIONSHIPS

Regarding these results, NRAs should carefully consider any intervention to regulate the MNOs-MVNOs relationships: the risk to obtain an "adverse" result in disturbing the infrastructure competition due to the weakening of the later entrants shall not be neglected. The following considerations deserved to be taken into account in that context to appropriately balance the regulation of wholesale markets where MNOs and MVNOs negotiate.

- ***Innovation in the mobile industry***

In the European mobile industry, it is still unclear if service-based competition may sustain more innovation while mobile users benefit from more choices on retail services coming both from the *"core"* or the *"edges"*.

Obviously, MVNO entries are likely to provide more innovation at the edge, for instance more game applications or music downloads. Innovation at the core includes change in technology platforms to provide better quality and new advanced services. With regard to mobile telephony, 3G technology provides better data transmission speed and service quality compared to GSM technology, and hence increases much consumer welfare.

From a practical perspective, there is no obvious documented evidence of the substantial benefits of innovation from service-based competition. It has not been proven that products and services delivered by MVNOs could not be provided by MNOs. Indeed, some MVNOs can be basically considered as resellers under the auspice of their brand names, as they are unable to supply innovative products or operate with *"lifestyle"* orientations.

- ***Irreversible investments and costs of delay***

As mentioned above, most infrastructure investment in the mobile industry can be regarded as irreversible which increases risks for MNOs as it is unlikely to recoup their investments in case of unfavourable market conditions to re-invest in other businesses. In the extreme case, improper regulation may lead to taking a no investment decision even when socially desirable (Sandbach 2006). In a more general case, MNOs may defer and take their investment decision when more information is gathered.

Although a late decision to invest might be beneficial for MNOs as they can employ more modern technology with lower costs, consumer and social welfare are greatly reduced. To give an order of quantification, we can mention Haussman (2000) who pointed out that the ten-year delay to introducing voice-messaging services might cost billions of dollars to US citizens.

In finding solutions for a proper methodology to regulate wholesale markets, real option analysis has proven to be useful in regulating wholesale tariffs. Much real option analysis in telecommunications has been developed and fundamentally shows that cost oriented wholesale tariffs do not provide investment incentives (Trigeorgis et al. 2000, Pindyck 2007). This is because MNOs bear all risks associated with infrastructure investments, and should be compensated with a mark-up on wholesale tariffs.

- ***Discrimination and net neutrality debate***

This topic has been controversial for both theorists and practitioners in telecommunications around the world. As discussed elsewhere in literature (Tirole 1990, for instance), some forms of discrimination can be Pareto optimal. In the mobile telephony context, "sellers" (MNOs) are somehow better informed than regulators over their "buyers" (MVNOs). For this reason, inappropriate intervention should constrain this ability and hence be detrimental to social welfare. Indeed, if innovations are great or if final retail products provided by an MVNO are highly differentiated, exclusory behaviour of the MNOs is unlikely to be the case since MNOs can increase their scale economy by supporting MVNOs.

Furthermore, European regulation principles have successfully overcome this concern by obliging transparency in wholesale offers without being able to discriminate, and implement vertical separation in telecommunications (Cave & Crocioni 2007).

- ***Cherry picking behaviour***

Broadly speaking, there is a high possibility that low access tariffs result in "cherry picking" activities. In particular, entrants could lure market shares by simply lowering retail prices without necessarily offering differentiated services. Put differently, regulating wholesale markets can be seen as an instrument to re-regulate mobile retail markets. As mentioned above, MNOs are better informed than regulators with respect to MVNOs' business. One way to prevent this opportunistic behaviour is to allow MNOs and MVNOs to bargain over access fees to obtain an optimal outcome without incurring regulatory costs.

- ***Ladder of investment***

Professor Martin Cave and his co-authors famously proposed an access mechanism in telecommunications that both encourage efficient entry and investments which is known as a *"ladder of investments".* Theoretically, new entrants can access incumbent networks at a low access fee so they can operate on the market. Access fees should then gradually increase as entrants establish their presence on the market and be able to invest in their own new networks. This mechanism can promote both service-based and infrastructure-based competition in the long run.

The ladder of investment theory, however, presents some limits. Some studies have pointed out the difficulty in implementing the "ladder of investment" in practice. Firstly, Oldale & Padilla (2004) argue that there may

be commitment problems: market players may not want to invest because of lower expected profits and wait to "free ride" rivals. Secondly, even when there is a regulatory commitment, it is still possible that the regulatory regime may penalise the *"good"*, and compensate the *"bad",* because later entrants have incentives to enjoy a longer beneficial regulatory period (Joao 2007). Later entrants do not compete very aggressively meaning that competitiveness may not be strengthened despite new market entry.

With respect to the mobile industry, Bourreau & al. (2008) show that it is impossible to apply the "ladder of investments" theory in mobile telephony because of social desirability, and especially due to limited spectrum frequency, limiting the possibility that a MVNO becomes a full MNO.

- ***Regulatory costs***

NRAs should also consider the regulatory cost that, even if sometimes hidden, can be high:

- Direct regulation costs, including employment costs, market analysis and any direct costs needed to make decisions.
- Indirect regulation costs can take place in many possible cases, including costs of taking "wrong" regulatory decision or giving the wrong signals and not provide sufficient incentives to market players.
- Lobbying costs: as there is regulation, operators may "lobby" regulators or try to get more favourable regulatory decisions.

CONCLUSION

It is important for European regulators to harmonise both types of competition in the telecommunication industry. One needs, however, to be cautious regarding intervention in mobile telephony because of its distinct characteristics, and not jeopardise current 3G network deployment. Furthermore, infrastructure-based competition in the mobile market has not shown an obivous "competitive failure" and it is believed that current technical convergence in telecommunications will further strengthen mobile industry competitiveness.

REFERENCES

Benzoni, L. (2007), "The "Curse of the Later Entrants: the Case of the European Mobile Markets". in Benzoni L., Geoffron P. (eds), "A Collection of Essays on Competition and Regulation with Asymmetries in Mobile Markets", Quantifica Publishing.

Benzoni L., Geoffron P. (eds), "A Collection of Essays on Competition and Regulation with Asymmetries in Mobile Markets", Quantifica Publishing.

Bergman M.A. (2004), "Competition in Services or Infrastructure based Competition?", Swedish Competition Authority and Stockholm University, September.

Bourreau M., Drouard J. & Feralli L. (2008), "The Ladder of Investment in the Mobile Market", in this issue.

Cave M. (2006), "Encouraging Infrastructure Competition via the Ladder of Investment", Telecommunications Policy, Vol. 30, pp. 223-237.

Cave M. (2004), "Making the Ladder of Investment Operational", Paper presented to the European Commission, November.

Cave M. (2006), "The Regulation of Access in Telecommunications: A European perspective", BEESLEY Lecture.

Cave M., Crocioni P. (2007), "Does Europe need Neutrality Rules?", International Journal of Communication, Vol. 1, pp 669-679.

Correa L. (2006), "The Economic Impact of Telecommunications Diffusion on UK Productivity Growth", Information Economics and Policy, Vol. 18, pp 385–404

Dispaso W., Lupi P., Manenti L. (2006), "Platform Competition and Broadband Uptake : Theory and Empirical Evidence from the European Union". Information Economics and Policy 18(1).

Dunnewijk T., Hultén S. (2007), "A Brief History of Mobile Communication in Europe". Telematics and Informatics, Vol. 24, pp 164-179.

European Commission (2007), "Progess Report on the Single European Electronic Communications Market (12th Report)"

European Commission (2008), "Progess Report on the Single European Electronic Communications Market (13th Report)"

Haussman J. (2000), "The Effects of Sunk Costs in Telecommunications Regulation". in Trigeorgis L., Hausman J., Tardiff T.(eds), The New Investment Theory of Real Options and its Implication for Telecommunications Economics, Springer US.

Henten A., Skouby K.E. (2006), "Service and/or Infrastructure Competition", Paper presented at 16th Biennial ITS conference.

João V., (2007), "Access Regulation under Asymmetric Information about Demand". FEUNL Working Paper No. 525 Available at SSRN: http://ssrn.com/abstract=1079967.

Jarspers F., Hulsink W. & Theeuwes J. (2005). "Virtual Enterprises, Mobile Markets and Volatile Customers – Entry and Innovation Strategies of Mobile Virtual Network Operators (MVNOs) into the Telecommunications Service Market", ERIM Report Series Research in Management.

Kittl J., Lundborg M., Ruhle E.O. (2006), "Infrastructure-based versus Service-based: Competition in Telecommunications", Communications and Strategies, No 64, pp 67-87.

Maiorano F., Stern J. (2007), "Institutions and Infrastructure Development in Low and Middle-Income Countries: The Case of Mobile Communications", Utilities Policy, No. 15, pp. 165-181.

Pindyck R. (2007), "Mandatory Unbundling and Irreversible Investments in Telecom Networks", Review of Network Economics, Vol. 6, Issue 3.

Schumpeter A.J. (1942), "Capitalism, Socialism and Democracy", New York: Harper & Row, 1942, 381 pp., Third edition, 1950.

Sandbach J. (2006), "Access and Termination Charges in Telecoms". IDEI Conference.

Tirole J. (1990), "The Theory of Industrial Organization", The MIT Press Cambridge.

Wallsten S. (2006), "Broadband and unbundling regulation in OECD countries", AEI Brooking Joint Centre on Regulation, Working Paper

Inter-Generational Transition in Technological Ecosystem : The Case of Mobile Telephony

SHAHZAD ANSARI
Department of Strategic Management and Business Environment RSM Erasmus University

RAGHU GARUD
The Pennsylvania State University

Abstract *The interregnum between successive technological generations has received little attention in technology studies. To understand what happens in the interregnum, we characterize technological change as happening within an ecosystem characterized by both momentum and inertia. Applying this framework to study the mobile communications ecosystem, we found that different parts of the ecosystem evolved at different rates with "collateral technologies" influencing the transition path that unfolded. We suggest that, rather than a distinct or unitary shift from an old to a new technology, transitions proceed in a zigzag manner resulting in the emergence of hybrid technologies.*

INTRODUCTION

Many studies on technological cycles, discontinuities, paradigms and trajectories are predicated on Schumpeter's (1934, 1942) analysis of creative destruction whereby waves of discontinuous technological change destroy old industries to create new ones (cf. Christensen 1997, Foster 1986, Hill & Rothaermel 2003, Tushman & Anderson 1986, Utterback 1994). Technological evolution is usually seen as proceeding in a sequential and progressive manner along an S-shaped curve[92]. The S-curve hypothesis suggests that the performance of a technology, slow at first, increases at a faster rate, finally flattening out to be supplanted by a new technology with its own S-curve.

The S-curve of technological evolution or the diffusion path of novel technologies (e.g. Griliches 1957) has been highly influential, especially in suggesting the point at which managers should shift investments from a mature technology to a new one. The S-curve has been used to depict the

92 *The S curve is has also been used in the innovation literature (Rogers 2003, Utterback 1994) to explain the diffusion rate of most innovations. Only a small number of people adopt an innovation at first, and then the adoption rate increases sharply, followed by a slowing of the laggards adopting the innovation.*

diffusion pattern of not just novel technologies and products, such as laser printing (Christensen 1997), process technologies (Karshenas & Stoneman 1993) and mini-mills (Tushman & Anderson 1986) but also a wide range of innovative social practices (e.g. Rogers 2003, Strang & Soule 1998).

Yet, those who have examined micro-processes involved in such transitions suggest that there is more than a simple linear shift from one technological system to another (Sood & Tellis 2004, 2005). Old technologies can prove surprisingly resilient and don't simply get eclipsed by new ones (Henderson, 1995). Instead, they may evolve through an "irregular step function" with big random improvements in performance, following long periods of dormancy (Sood & Tellis 2004). Also, major advances within a technology regime can also drive the evolution of technologies (Lawless & Anderson 1996). Moreover, different interrelated constituents of the socio-technical ecosystem evolve at different rates (Adner 2006, Hughes 1987, MacKenzie 1987) mediated by bottlenecks (Rosenberg 1982) and "reverse salients" (Hughes 1983: 73). In sum, the transition from an old technology to a new one is neither inevitable nor sequential.

What are the processes that may characterize inter-generational transitions? We address this question by analyzing the mobile communications industry that has often been marked by technological shifts. In particular, we study the dynamics in the *interim* period between the second and third generation technologies. To do so, we adopt a perspective that allows us to examine the diverse pulls and pushes created by constituents. Our study suggests that, rather than the "winner takes all" tipping effects, or the inertia caused by "lockouts" and high switching costs (David 1985, Schilling 2003), technological transitions may be characterized by both momentum and inertia. As a result, technological transitions are not necessarily *unitary* shifts from the old to the new, but, instead may be characterized by uneven movements that generate asynchronies or imbalances across the ecosystem during the migration process. By highlighting these processes, we offer a more multifaceted conceptualization of technological transitions that includes the institutional and ecological dynamics and carries important caveats for premature jumps to the next technological phase.

How do these contradictory forces of momentum and inertia play out during technological transitions? Do constituents of an ecosystem operate in sync or are there temporal discords that emerge during various points of a technological transition? And, how do these processes influence migration paths? We address these questions by providing an in-depth account of the processes involved in the transition between the second and third generation mobile telecommunication technology. But, first, we offer a brief overview of our research site.

TECHNOGICAL EVOLUTION IN MOBILE COMMUNICATIONS

We analyzed the mobile communications ecosystem between 1999 and 2005. On April 19, 2000, the UK government held the first 3G spectrum auctions in Europe, raising an unprecedented $35 billion from the sale of five licenses. Soon thereafter, the spectrum auction in Germany raised almost $50 billion. While subsequent auction proceeds from spectrum auctions in Europe did not reach such surreal levels, the combined revenues from the sale of licenses for 3G mobile technologies raised more than $100 billion as many mobile operators made their biggest ever investments. The timing of the auctions had coincided with the peak of the dotcom boom that had not only made these investments seem justifiable at the time but had also allowed access to 'easy money' from willing lenders.

Five years down the road, the demand for products and services based on 3G technologies remained well below expectations. Many were convinced that the enormous prices paid to obtain these licenses were completely unrealistic as far as the potential for future profits was concerned, and that firms were saddled with the winner's curse (Kagel & Levin 1986). Indeed, even just after their completion, the auctions were variously described as the 'wireless gamble' (The Economist 2000) and a 'spectrum land grab' for a new and untried technology. Many mobile operators delayed rolling out infrastructure for 3G systems and wrote down the value of their spectrum licenses while some even returned these to the state at considerable loss. Despite the tremendous optimism that the third generation mobile technologies had generated at the turn of the millennium, lack of interest from end users kept penetration rates extremely low. Even by 2005, out of the almost 2 billion mobile customers worldwide, just about 2% had migrated to 3G systems and even those customers sparingly used the various 3G services. See next table for subscriber numbers for the various technological generations.

Table 5. Millions of subscribers on 2G and 3G systems

	2002	2003	2004	2005
World	1137.8	1382.9	1714.1	2177.1
GSM (2G)	809.3	1012.0	1296.0	1709.2
Others (1G and 2G and 2.5G)	328.3	368.1	401.3	417.9
3GSM (3G)	0.2	2.8	16.3	50

In the wake of such low demand for the new technology, many operators decided to focus on upgrading the previous generation 2G systems to what came to be known as 2.5G. This interim 'solution' provided almost all of the benefits that 3G had promised, yet at a fraction of the cost and without the need for costly new spectrum licenses. As the label 2.5G suggests, instead of a smooth transition from the second to the third generation, the system

settled down somewhere in between, incorporating facets from both generations. To understand how and why this happened, we need to look far beyond the technologies at play. A richer understanding would entail studying the various constituents of the entire ecosystem in which the technologies are embedded (Adner 2006, Van de Ven & Garud 1993, Geels 2002). We describe such an ecosystem for the mobile communications industry.

- ***The Mobile Ecosystem***

At the time of the auction, the mobile ecosystem, including handsets and other co-specialized assets, was locked into a 2G mode that had reached saturation levels of penetration. For a new technological system like 3G to work, it would require not only new core and complementary technologies but also several co-specialized assets (Teece 1986). These assets include, for instance, new generation handsets from manufacturers (that are also compatible with previous generation mobile phones), new base stations and masts from infrastructure providers for the transmission of 3G signals and compelling 'content' (video games, websites etc. for mobile handsets) from application developers. As the competences of mobile firms operating under 2G were limited to providing voice-based services, they had to enroll a wide range of stakeholders, such as banks and entertainment companies to develop new kinds of services – mobile banking and purchasing, interactive video games on the mobile etc.

Since spectrum is a state-controlled resource, the ecosystem also involved regulators. Similarly, infomediaries – analysts and media – also played an important role during the transition process. Besides, environmental groups concerned about the potential radiation hazards from 3G systems were important social groups influencing the transition process[93]. This social group urged municipalities to resist allotting building permits for new sites for 3G transmitters. Finally, any technological ecosystem will remain isolated without a critical mass of users who attach various meanings to products, vis-à-vis their value in use. For 3G to become viable, end users would have to show an interest in using 3G services. Before we elaborate further on the roles of various actors in the mobile ecosystem, we first provide a brief overview of the essential characteristics of 3G technologies and the advantages it promised over the preceding 2G systems.

93 *3G systems operate at higher frequencies that required the construction of additional masts and base stations in cities as well as more powerful handsets than 2G technologies. Many believe that these components posed significant radiation hazards.*

➢ *Key Characteristics of the Third Generation (3G) Wireless Technology*

As against the globally dominant GSM (Global System Mobile) standard in 2G technologies, 3G technologies used the CDMA (Code division multiple access) radio interface standard. CDMA is far more efficient in 'farming' available spectrum space by assigning a unique *code* for each frequency channel, thereby providing more traffic per megahertz of spectrum (Funk & Methe 2001). Most 2G technologies, including the European-led GSM standard, was based on TDMA (Time division multiple access) – a technology that creates multiple access channels for subscribers by *spacing* frequencies in time.[94] 3G systems not only made more efficient use of spectrum but also had the capability to transmit large amount of data at much higher speeds by using a bigger band of spectrum as compared to 2G. As against 14.4 kilo bytes per second (kbps) 3G could transmit data at up to 2000 kbps.

Besides these differences in the core technologies surrounding the two generations, 3G also offered a more efficient way of transmitting wireless data. All third generation technologies use 'packet based' systems for sending data – a technology employed in the Internet. At the time of originating a data file's transmission, the file is split up into smaller units, or packets, containing identifiable information, which is needed for their re-assembly at the file's destination. Available bandwidth usage is optimized by minimizing the transmission 'latency' (i.e. the time it takes for data to pass across the network). In contrast, 2G technologies use circuit-switched technology that sets up a dedicated connection between two callers for the entire duration of the communication. Even if no actual communication is taking place, the channel still remains unavailable to other users, creating inefficiencies in spectrum utilization.

While the core 3G technology, CDMA, along with the use of packet switching for data transfers allow better utilization of spectrum, the need to handle large amounts of data also need higher spectrum capacity or bandwidth. Spectrum being a valuable resource meant that operators had to pay billions of dollars just for obtaining spectrum licenses for 3G systems, let alone build technological infrastructure and market the new technology. To understand the exorbitant prices paid for 3G licenses during many of the auctions, we need to look at the antecedents to the 3G auctions. This includes the context in which the auctions took place and the activities of various constituents in the run up to the auctions.

---

94 One exception was the use of CDMA in incumbent 2G systems, such as the CDMA-One (IS-95) that used 2G spectrum. However, with the TDMA-based GSM being the de facto global standard for 2G, the use of CDMA-One remained largely limited to North America (Funk & Methe 2001).

➢ *The 3G Spectrum Auction*

Collectively, the interrelated constituents and the various social groups comprising the mobile ecosystem generated tremendous optimism for the new technology, just prior to the spectrum auctions. It was no wonder that the industry leapt at the chance to get a piece of the spectrum when the governments designated a specific bandwidth and put up licenses for sale in 2000-2001. Overall, European operators invested over 110 billion Euros in what was one of the biggest investments in recorded history by any industry on the introduction of a new technology[95]. In comparison, the license fee from 1994-2002 for 2G spectrum was in the range of just 10 billion Euros for the entire European market. In the case of 3G, just in the UK auction for instance – the first to be held in Europe raised about $35 billion from selling 5 licenses – 4,000% higher than the minimum price set by the UK government (Borgers & Dustmann 2001)[96]. The UK auction recorded the highest revenues in part because of its Anglo-Dutch design – a hybrid – that could capture the virtues of both the ascending (English) and sealed-bid (Dutch) forms (Klemperer, 2002). In this design, the auctioneer begins by running an ascending auction until just two bidders are willing to pay the current price. The two are then each required to make a 'best and final' sealed-bid offer and the winner pays his bid. This format reduces the incentives for firms to form consortia prior to the bidding, as they can be beaten in the final round. It also makes tacit collusion harder, because the sealed-bid stage allows firms to renege on any tacit deals without fear of retaliation.

The high prices were paid despite the stringent rollout obligations attached to spectrum licenses. For instance, the UK regulators imposed rollout obligations with the condition that 3G services by each licensee were to cover at least 80% of the UK population by the end of 2007 (Ofcom 2000). As 3G required a completely new infrastructure, meeting this obligation in

---

95 While auctioning spectra licenses was unprecedented in Europe, the regulators in charge of 3G licenses had taken a cue from the spectacular success of spectrum auctions in the US. The US regulator, Federal Communications Commission (FCC) had raised an astonishing $42 billion between 1994 and 2000, described as the "49 California gold rush" (Business Week, 1993). Also, some countries such as Japan and Sweden had decided to stick with the format of using administrative mechanisms (beauty contests) for allocating 3G licenses, as had been the case during the previous generations of mobile telephony, when these licenses were simply awarded to 'deserving' firms.

96 While almost every incumbent operator bid for a 3G licenses, a notable exception was the French firm, Bouygues. Its CEO, despite intense pressures to make a switch to 3G chose not to apply for a license. He was later vindicated when the French government slashed license fees and the operator obtained a license at less than a quarter of the original license fee (Baker & Clifford, 2002).

sparse demographic distributions and terrain (providing coverage to the 'last mile') would require huge further investments in addition to the payment for licenses. The total projected investment for rolling out 3G network infrastructure from 2001-2010 in Europe was about another 250 billion Euros. Mobile firms nevertheless made huge investments in their commitment to 3G despite these tough license requirements.

- ***The Auction Aftermath***

More than five years down the road, the migration had not occurred as intended despite a concerted effort at 'technological push' from many operators. Analysts forecast at the time of the auctions had been that the technology would be up and running by 2003-2004 and generate significant rents by 2005. Even at the end of 2005, 3G penetration levels remained almost insignificant. In terms of market share, out of the almost 2 billion mobile customers worldwide, just about 2% were on 3G systems and even these customers sparingly used the various 3G services. It still remained unclear which type of 3G data services would utilize the excess capacity and be popular with end-users.

Given such low diffusion rates, an increasing number of mobile firms had to scale back their targets, withdraw from unattractive markets, and delay or even abandon 3G rollouts. The enormous sums spent at the auctions, now seemed an extremely unwise investment. The Financial Times summed up the situation in these words: "Taking into account the write-offs, bankruptcies and closures worldwide, probably $1,000 billion has gone up in smoke" (Financial Times 2001). With customers evincing little interest, and most analysts and investors increasingly skeptical about 3G, the future of the technology looked far from bright. Even some of the mobile operators that had invested billions began to express doubts about the economic viability of 3G. This was evident when the chief executive of BT Cellnet, Peter Erskine stated: "There is recognition that too much was paid for 3G licenses."

- ***Extension of the Old Technology***

2G was upgraded to what came to be known as two-and-a-half generation (2.5G) or GPRS (General Packet Radio Services) through bolting on packet-switching data transmission technology onto 2G systems. Sometimes called '3G-lite' in the industry, these 2.5G upgrades essentially allowed faster connections to the Internet via the mobile phone. 2.5G could handle data transmission speeds between 33.6 and 128 Kbps, as against a maximum of 2000kbps for 3G, thereby lifting the performance to a level which is in the middle range of 2G and 3G. Thus, while 2.5G marked a significant improvement in 2G systems, it nevertheless operated on the limited 2G

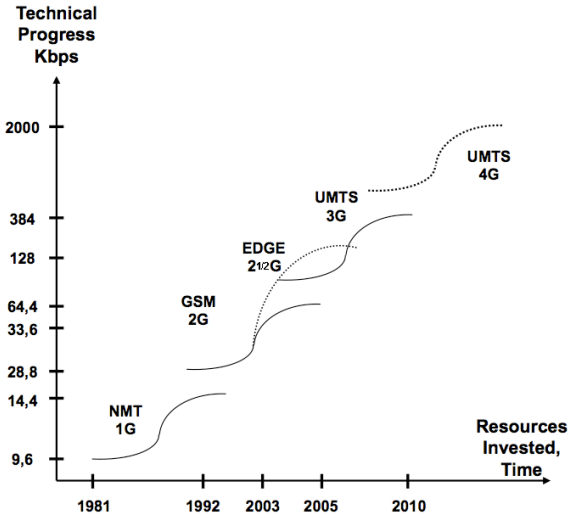
spectrum and did not achieve what 3G was capable of. As an example; downloading a song with a 2.5G network would take over a minute as against 5 seconds on a 3G network.

As the mobile ecosystem seemed to temporarily settle down around 2.5G technologies, a number of mobile firms with 3G licenses, such as Vodafone, delayed their 3G launches beyond initial promises and focused on upgrading their 2G networks to 2.5G. Firms without 3G licenses, such as the Danish operators, Sonofon and Telias also decided to focus on 2.5G and offer most of what were supposed to be '3G' services. Many operators now acknowledged that the much-hyped transition to 3G was unlikely to be momentous and began to regard 3G as *evolutionary* instead of revolutionary. Indeed, as Dave McGlade, CEO of BT Cellnet (later known as mmO_2) stated in 2003: "In terms of the kinds of applications our customers are going to use, I don't think there is a huge difference between 2.5G and 3G" (Telecom Review – Conference proceedings in the 3GSM World Congress in 2003).

While many operators chose the 2.5G route, they faced a dilemma – preventing the cannibalization of their customer base network in a near-saturated market while, at the same time, also requiring them to sign up for new expensive 3G contracts to increase the ARPU. The situation actually was far more complex than having to decide between two rival basic technologies. According to Bob Merritt, VP of Semico Research Corp. "It is going to be a case of supporting 2G, 2.5G and 3G, all at the same time. This isn't going to be a nice, smooth transition."

Worse still, was the prospect that firms that chose to upgrade to 2.5G could bypass 3G and make a direct leap for 4G (The Economist, 2003). These 4G system were expected to include new technologies such as improved modulation that could deliver full motion video on mobile devices. A former high ranking objective of a major mobile operator we interviewed suggested that 3G could well be a "stillborn" technology and the third generation a "lost generation" (Former head of technology of NTL, the second largest telecom in UK). Thus, it was conceivable that 3G could simply be leapfrogged by what was being labelled as 4G. The next figure depicts the development of performance for mobile technologies from 1G to 4G (projected) where each technology is introduced at a higher performance level than the former technology.

Figure 19. Neither fish nor fowl but 2 ½ G



2G was upgraded to what came to be known as two-and-a-half generation (2.5G) or GPRS (General Packet Radio Services) through bolting on packet-switching data transmission technology onto 2G systems. Sometimes called '3G-lite' in the industry, these 2.5G upgrades essentially allowed faster connections to the Internet via the mobile phone. 2.5G could handle data transmission speeds between 33.6 and 128 Kbps, as against a maximum of 2000kbps for 3G, thereby lifting the performance to a higher level than 2G but well below 3G's performance. As 2.5G operated on the limited 2G spectrum, it could not achieve what 3G was capable of. Nevertheless it was able to close the gap in performance between the two platforms thereby generating increasing inertia to a move to 3G.

DISCUSSION

The body of work on technological transitions has done little to address real-time transition issues at play, such as the forces that arise as a result of interactions between producers, users, and institutional players constituting a technology ecosystem. To develop a richer understanding of the dynamics of technological transitions, we examined recent developments in mobile telephony, focusing in particular, on the interregnum between two successive technological generations, 2G and 3G. As we have observed, the process of technological transition can be messy, erratic and contested, not necessarily driven by smooth shifts; aspects of the past technologies continue to matter; and new technologies with higher functionalities do not simply eliminate the older ones in a process of creative destruction.

In the light of our study of mobile telephony, we have argued for the need to conceptualize networks as a complex ecosystem in order to develop a richer understanding of how transitions unfold. We now draw on our findings to develop a richer imagery of technological generations.

➢ *Towards a richer understanding of technological transitions*

While scholars have discussed the need to take an expanded network or systemic perspective in conceptualizing technological transitions, the various elements of the system are usually seen to be operating in sync as it moves forward from one generation to another (e.g. Hughes 1983). However, by disaggregating the elements of an ecosystem, we found that, rather than in being sync, the different elements evolved at different rates, thereby setting up the transition to temporal discords. It is understandable that a delay in co-specialized and complementary assets makes it more likely that old technology would catch up with the new technology. These improvements may come from within the industry but may also result from developments in related industries. For instance, in the context of mobile communications, a "collateral technology" (the bolting of packet switching onto existing 2G technology) imported from the domain of the Internet appears to have been the true 'radical shift' at a time when most constituents were preoccupied with core 3G technologies. By drawing on elements from the new technologies, this collateral technology created links between different industries – Internet and 2G mobile – and enabled the extension of the previous generation (2G). According to most industry experts, the collateral technology represented a bigger ground shift in mobile communications than core 3G technologies, despite the massive investments that went into developing 3G. Indeed, users did not find 3G to be dramatically different from 2G and most seemed satisfied with upgrades of the old technology that served almost all of their needs. Thus, radical technological advances are not necessarily 'disruptive' for all the constituents of an ecosystem.

The phenomenon of how collateral technologies can revitalize previous generation technologies has been observed in other domains. For instance, improvements in steelmaking from the introduction of open-hearth furnaces in the late 1870's enabled the substitution of earlier generation steam boats for sailing ships. Open hearth furnaces allowed the production of better steel, which in turn enabled boiler plates and boiler tubes to withstand higher pressures. As a result, more efficient steam boats could be operated profitably (Dattee 2007). Another example of a collateral technology to extend the life of an old technology can be seen with photo-lithography where optical lithography has shown unusual persistence as the dominant manufacturing technology for computer chips since the late 1960s (Henderson 1995). While the industry was focused on developing "next generation lithography" (NGL) that included technologies such as extreme

ultraviolet lithography (EUVL), the discovery of "enhanced liquid immersion" extended the life of optical lithography[97] (Sydow & al. 2007). Infusing optical lithography with a collateral technology required only minor adjustments compared to EUVL for not only extending the path of optical technology but also substantially improving its performance. In short, collateral innovations in related industries can result in extending the life of the old technology and close its performance gap, vis-à-vis the new technology. Thus, technological developments in related industries also matter in processes of technological transitions.

CONCLUSION

Gaining a deeper understanding of the processes and dynamics whereby transitions between generations occur holds important implications for policy and strategy. In this regard, our paper offers important insights. First, by conceptualizing the journey as involving a multiplicity of heterogeneous elements, we expand the scope of inquiry for the sources of change driving the process. Sponsoring a novel technology involves large upfront investments that can drastically transform the social, economic, and organizational landscapes. With so much at stake, a thorough understanding of the role of multiple constituents and how collateral technologies mediate outcomes in transition dynamics is crucial to firms and policy-makers. Technological evolutions, thus, are not determined simply by a battle between competing technologies. Managers making technology investment decisions need to also pay attention to the interdependencies among various heterogeneous elements in a technology ecosystem, the inter-relatedness and spillovers that may occur among various technologies and the institutional dynamics that surround particular technologies.

Second, our study shows that there is value in disaggregating these elements and in exploring the differential rates at which these elements develop. Such a disaggregation highlights potential temporal discords between the different elements. As the 2G/3G example shows, how these temporal discords play out has a profound impact on the rate and direction of change of the overall system. Specifically, the differential rates at which the new core and co-specialized assets emerge as well as the rates at which existing technologies improve, fueled by speculations by infomediaries such as the media, are parameters to watch out for during the transition process.

97 *In enhanced liquid immersion lithography (LIL), a drop of fluid (water or oil) is placed between the optical lens and the wafer. Compared to air, the higher refractive index of fluids leads to a better image resolution. Originally used in microscopy to enlarge the image of the specimen, immersion is used in optical lithography to print miniaturized features onto silicon wafers.*

This observation places a word of caution against unrelenting belief in the supremacy of a novel technology that can render their sponsors "blind with respect to other technological possibilities" (Dosi 1982), and the prevailing wisdom to abandon old technologies that are sometimes perceived as having reached their natural performance limits. It thus calls for more attention to parallel and complementary developments in both the technical and institutional environments that impact technological transitions.

Future studies in technological evolution should examine multifaceted and complex technological ecosystems in other contexts to gain a better understanding of inter-generational transitions and the dynamics of the intervening period. What are the antecedents of momentum that causes premature jumps to a new technology and what generates inertial forces that stall the migration? When and why do various constituents of the ecosystem evolve at variable rates and what causes them to move in sync? Both fine-grained qualitative accounts as well as large sample quantitative studies can lead to a better theory of technological transitions.

REFERENCES

Adner R. (2006), "Match Your Innovation Strategy to Your Innovation Ecosystem", Harvard Business Review, 84(4), 98-107.

Anderson P., Tushman M.L. (1990), "Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change", Administrative Science Quarterly, 35, 604-633.

Baker S., Clifford M. (2002), Tale of a Bubble How the 3G fiasco came close to wrecking Europe. Business Week, June 03.

Borgers T, Dustmann C. (2003), "Awarding Telecom Licenses: the Recent European Experience", Economic Policy, 18(36), 215.

Christensen C.M. (1997), "The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail", Cambridge, MA, HBS Press.

Dattee B. (2007), "Appropriability, Proximity, Routines and Innovation. Challenging the S-Curve: Patterns of Technological Substitution", Paper presented at the DRUID Summer Conference, Copenhagen, CBS, Denmark, June, 18 – 20.

David P.A.(1985), "Clio and the Economics of QWERTY", American Economic Review, 75(2).

Dosi G. (1982), "Technological paradigms and technological trajectories", Research Policy, 11, 147-162.

Financial Times (1999), "Survey, FT Telecoms: '3G Technology Will Transform Internet Access: Viewpoint'", November 24.

Foster R. (1986), "Innovation: The Attacker's Advantage", New York: Summit Books.

Funk J.L., Methe D.T. (2001), "Market and Committee-Based Mechanisms in the Creation and Diffusion of Global Industry Standards: The Case of Mobile Communication", Research Policy, Vol. 30, pp. 589-610.

Geels F.W. (2002), "Technological Transitions And System Innovations: A Co-evolutionary And Socio-technical Analysis", Blackwell Publishing.

Griliches Z. (1957), "Hybrid Corn: an Exploration in the Economics of Technological Change", Econometrica, 25(4): 501.

Henderson R. (1995), "Of Life Cycles Real and Imaginary: The Unexpected Long Old Age of Optical Lithography", Research Policy, 24 (4): 631-644.

Hill C.W.L., Rothaermel F.T. (2003), "The performance of incumbent firms in the face of radical technology innovation", Academy of Management Review, Vol. 28:2, p. 257.

Hughes T. (1983), "Networks of Power", Baltimore, MD & London: Johns Hopkins University Press.

Kagel J.H., Levin D. (1986), "The Winner's Curse and Public Value Information in Private Auctions", American Economic Review, 76 (5): 894-920.

Karshenas M., Stoneman P.L. (1993), "Rank, Stock, Order, and Epidemic Effects in the Diffusion of New Process Technologies - an Empirical-Model", Rand Journal of Economics, 24(4): 503-528

Katz M., Shapiro C. (1985), "Network Externalities, Competition and Compatibility", American Economic Review, 75 (3): 424-440.

Klemperer P. (2002), "How (not) to Run Auctions: The European 3G Telecom Auctions", European Economic Review, Vol. 46 Issue 4/5, p829-845.

Lawless M. W., Anderson P. C. (1996), "Generational Technological Change: Effects of Innovation and Local Rivalry on Performance", Academy of Management Journal 39(5), 1185-1217.

Levinthal D.A. (1998), "The Slow Pace of Rapid Technological Change: Gradualism and Punctuation in Technological Change", Industrial & Corporate Change, Vol. 7 Issue 2, 217-47.

MacKenzie D. (1987), "Missile Accuracy: A Case Study in the Social Processes of Technological Change", pp. 195–222 in Bijker T. Hughes T. & Pinch T. (eds), "The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology", Cambridge, Mass. MIT Press.

Rogers E.M. (2003), "The Diffusion of Innovations", 5th Edition. New York, Free Press.

Rosenberg N. (1982), "Inside the Blackbox. Technology and Economics", Cambridge University Press, New York.

Schilling M.A. (2003), "Technological Leapfrogging: Lessons from the U.S. Videogame Industry", California Management Review, 45(3):6-32.

Schumpeter J.A. (1942), "Capitalism, Socialism and Democracy", George Allen & Unwin London, 1976.

Schumpeter J.A. (1934), "The Theory of Economic Development", Cambridge, MA: Harvard University Press.

Sood A., Tellis G.J. (2004), "Technological Evolution and Radical Innovation", Journal of Marketing, Vol. 69, 152–168.

Sood A, Tellis G.J. (2005), "The S-Curve of Technological Evolution: Strategic Law or Self-Fulfilling Prophecy", Working Paper No. 04-116. Boston, Marketing Science Institute.

Strang D., Soule S.A. (1998), "Diffusion in Organizations and Social Movements: from Hybrid Corn to Poison Pills", Annual Review of Sociology, 24, 265-290.

Sydow J., Windeler A., Schubert C., Möllering, G. (2007), "Organizing Networks for Path Creation and Extension in Semiconductor Manufacturing Technologies", Paper presented at the Academy of Management Meeting, Philadelphia, August 8th.

Teece D.J. (1986), "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy", Research Policy, Vol. 15, No 6, pp 285-305.

The Economist (2000), "Europe's Mobile Telephony: Waves of the Future", July 8th.

The Economist (2003), "Wireless Telecoms. Move over 3G. Here comes 4G", May 29th.

The Financial Times (2001), "Telecoms Crash", September 5th.

Tushman M.L., Anderson P. (1986), "Technological Discontinuities and Organizational Environments", Administrative Science Quarterly, 31, 439-65.

Utterback J. (1994), "Mastering the Dynamics of Innovation, Cambridge", Harvard Business School Press.

Van de Ven A.H., Garud R. (1993), "Innovation and Industry Development: The Case of Cochlear Implants.", in Burgelman R., Rosenbloom R. (eds.), "Research on Technological Innovation Management and Policy", 5: 1-46. Greenwich. CT: AI Press

Mobile Virtual Network Operators: A Strategic Transaction Cost Analysis

SVEIN ULSET
Norwegian School of Economics and Business Administration, Bergen

***Abstract**: This paper describes and analyzes critical conditions for achieving net benefit from opening the value-chain in mobile communications by introducing mobile virtual network operators (MVNOs). MVNOs are radio-less network operators that outsource the radio part and some other network elements to radio-based mobile network operators. Preliminary experience from the early development phase suggest that MVNOs offering complex bundles of innovative value-added services will not be competitively sustainable as separate firms, only as more tightly integrated partners of radio-based mobile network operators. As the MVNO market develop into more mature phases, supported by more entry-friendly regulatory, technical and contractual practices, the number of competitively sustainable MVNOs will increase along with the scope of value-added services offered by these specialized firms.*

INTRODUCTION

The purpose of this paper is to evaluate the economic benefit from the joint operation of *regular* and *virtual* mobile network operators through the lenses of strategic transaction cost economics (Williamson, 1999a). Mobile virtual network operators (MVNOs) are *radio-less* operators that own and control at least some part of the mobile network, while contracting out to regular *radio-based* mobile network operators (MNOs) the radio part and any remaining complementary network facilities and service applications that are necessary to provide mobile services to end users[98]. Prospective benefits are sharper upstream competition, lower supply prices and more innovative downstream service packaging, branding and marketing. These are benefits from

98 *Besides owning a minimum of network elements to deliver and receive calls, control over a Subscriber Identity Module (SIM) card is recognized as a key requirement of an MVNO. Sufficient user "control" may, however, be achieved without legally owning the card's identification code, including the mobile network code that would allows the MVNO subscribers to roam into the network of various MNOs. This can be achieved by contracting with the legal owner of the card code for the right to use the card for all or almost all commercial purposes. Although Sense uses their own SIM Toolkit and may also claim they sell their own SIM cards, it is still Telenor or NetCom as spectrum-licensed network operators that are the legal holder or "owner" of the mobile network code. In this paper service providers of the Sense-type are included as a kind of virtual operator because they represent valuable experience.*

outsourcing that seldom can be achieved without additional *transaction costs* in terms of time-consuming negotiations, arbitration and even litigations. Net social benefit can therefore only be accomplished to the degree transaction hazards are moderate throughout the contracting period, or to the degree contractual safeguards are sufficient to prevent transaction costs from escalating.

Until recently the prevailing opinion of incumbent operators and most regulators has been that the positive effects of virtual network operation were highly uncertain, probably minimal and definitely not big high enough to justify regulatory intervention[99]. The main reason for such skepticism seems to be the expected negative incentive effects that such a regulation would have on investment in future networks such as in 3rd generation mobile network (UMTS network), combined with weaker infrastructure competition. Although these negative investment incentive and competition effects will be moderated by the positive effect that such regulation will have on service innovations and competition in service provision, skeptics (most incumbents and some regulators) expect positive effects to be outweighed by negative ones. In particular, since the radio-based network owners stand to benefit less than others, virtual operation will not appear as a viable strategy until facilitating regulation and appropriate operating conditions have been established, the associated industrial dynamics sufficiently documented, and the radio-based network operators highly convinced about their own benefit from such dynamics.

My intention with this paper is to start a more open and critical discussion and evaluation of this difficult, but important question. Our choice of theoretical approach for such an evaluation will be presented in section 2. Further definition and description of the concept of virtual mobile network operators are covered in section 3, followed by short accounts of early and recent developments in section 4 and 5. Conclusions finalize the paper in section 6.

A STRATEGIC TRANSACTION COST ECONOMICS APPROACH

Transaction cost economics (TCE) deals with transaction hazards caused by *interdependency* and *asymmetric information* and the respective *governance* structures (firm, markets, hybrid contracting) that may serve to mitigate such hazards. TCE can also be applied strategically to explore how different

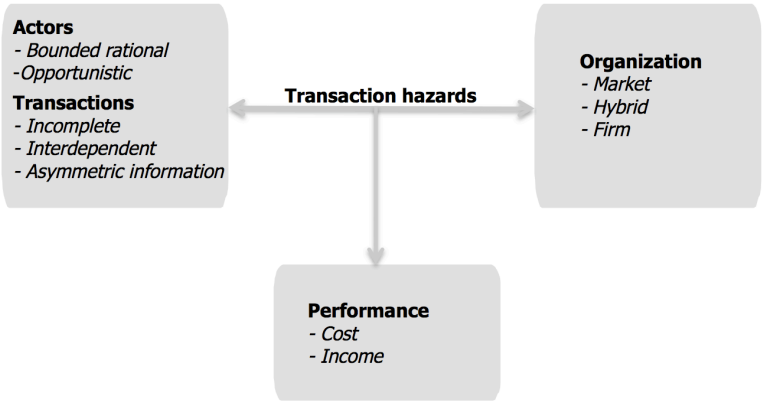
99 *This was the conclusion reached by the British regulator Oftel and Norwegian Department of Communication that recently evaluated the concept of virtual operators in mobile communication (Oftel 1999a and 1999b, St.meld.nr. 24 (1999-2000)).*

governance forms may assist in exploiting competitive advantage that can be derived from leading technology and best practice. As indicated above, the crucial MVNO question is whether network services outsourced to radio-based operators is *technologically separable* from complementary services provided by virtual operators, and if separable, whether they still are too interdependent (and therefore *non-redeployable*) to justify full *corporate separation*.

Technological separability will increase and interdependency will decrease to the degree the respective inter-firm transactions between regular and virtual operators are facilitated by open *interface standards*. For our purpose, the latter will include standard operating procedures not only for technical but also for human interaction. Such operating procedures facilitate interoperability between interactive facilities, programs and participating actors by providing a common "platform" consisting not only of shared language, technology and expertise, but also of shared values and norms. As a consequence, inter-firm communication will be improved and transaction costs will decline.

Remaining hazards are mitigated by selecting among governance forms (firm, market, hybrid forms) characterized by different attributes (incentive, control, contract laws) in accordance with the following recipe: *Incomplete and interdependent transactions carried out by bounded rational and opportunistic players should be organized in firms or under firm-like (hybrid) contracts rather than under market contracts*. Compared to firms, markets are characterized by stronger *economic incentives*, weaker *administrative control* and more reliance on the *court system* as conflict resolution mechanism (Williamson 1991). Hybrid contracts such as joint ventures or long-term exclusive contracts are defined by intermediate value combinations.

Figure 20. The Core Issue



That is, radio-less "virtual" network operators that buy basic network services from radio-based "non-virtual" operators will normally do this under incomplete contracts that specify some, but not all the obligations under some, but not all future conditions. Instead of specifying every possible future decision and condition, difficulties are dealt with, and conflicts resolved, as the future unfolds. This may work reasonably well under most ordinary supply contracts as long as problems are simple and easy to solve and potential losses from switching partner are small. In situations where problems are more difficult to solve, and switching costs are large, simple contracts will no longer suffice. Should one of the parties, due to possible failures or defects by the other party wish to exit from the relation, this may not only lead to time-consuming and costly conflicts, arbitration and possible litigation, but also to the loss of all the non-redeployable assets. To avoid such transaction costs, simple contracts should be replaced by more complex contracts with stronger *safeguards*, such as long-term contracts, joint ventures or fully integrated corporations, dependent on the level of contractual difficulty and the size of potential losses from separation. The choice of integrated corporation should consequently be reserved for the most difficult cases with the highest loss potentials[100].

However even under highly standardized conditions *ambiguity* and *asymmetric information* may still cause substantial transaction costs. Due to the subjectivity involved in allocating fixed cost to the use of unbundled network elements, correct cost-based pricing of network element usage has been very difficult, almost impossible. As a consequence, conflicts over pricing issues have been frequent, and litigations part of everyday life.

100 This distinction is of course similar to the distinction between autonomous and systemic innovation (Teece 1986, Chesbrough and Teece 1996).

Should the infrastructure be forced open by regulators, incumbents would naturally fear that revenue from the sales of network capacity would no longer cover total network costs, thus destroying their incentives for investing in new infrastructure and in developing new services. Even though the regulator officially prescribes and expects that the same quality and price should be offered external customers as internal units, this objective may never be fully attained, due to the many indirect ways vertically integrated operators may apply to reimburse and favour their own service providers. In turbulent times when technology standards, regulations and business practices are still evolving, and business partners behave opportunistically, causal ambiguity and asymmetric information will work in the disfavour of external service providers.

Nevertheless, the growing number of separate resellers and virtual operators, especially in the more advances markets, suggests the opposite trend that vertical integrated corporations are increasingly being replaced or supplemented with hybrid or market contracting. Apparently, not only primary network services but also associated support and customer care have become increasingly tradable, primarily due to digitization and standardization of network technology. Software programs that interoperate across standard interfaces, increasingly handle transactions previously handled by humans.

From a strategic point of view, most appropriate governance mechanisms should be applied (in accordance with the above recipe) to exploit most promising sources of competitive advantage. In particular, virtual as well as non-virtual operators may achieve extra profit by specializing on a narrower field of resources to the degree these resources are *valuable*, *rare*, *inimitable* and *well organized* (Barney 1997, chap. 5). Resources such as physical and human assets are valuable to the degree their positive effects on profit or value added are large. They are rare in the sense of being *scarce* (few competing suppliers), *different* (rather unique compared to similar assets at the competitors' disposal), *non-substitutable* (no alternative asset or technology to perform the function; natural monopoly) or *specific* (less productive value for alternative users or in alternative uses). Being rare in the absolute sense means that no useful substitute exists (e.g. telephone lines before cellular and satellite) in which case the network operator may charge *monopoly price* on his services.

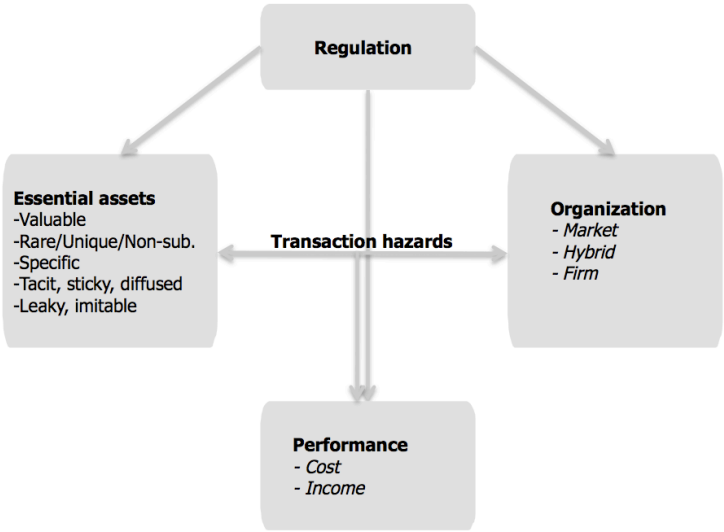
Furthermore, resources such as technology, knowledge or competence may be difficult to imitate or copy to the degree they are invisible, tacit, sticky or diffused. Being easily imitable means that monopoly rent from unique products will be competed away rather quickly. Should patents be unavailable or insufficient, internal organization will at least give some protection against early leakage, particularly if strengthen with measures such as co-ownership, deferred payment, confidentiality clauses, and socially conditioned probity (Liebeskind 1995, Williamson 1999b). Leading

technology and best practice may therefore be considered "well organized" to the degree they are safeguarded with these and similar firm-like mechanisms against the hazards of leakage, opportunistic recontracting, expropriation and improper usage. Conversely, as technical interfaces are standardized, interdependencies are dissolved and early technology leakage eliminated, hybrid and arm's length contracting might gradually replace integrated firms (fully integrated radio-based operators).

In the following figure the arrows linking organization, essential assets and performance summarize these relationships. To the degree uniquely valuable assets (leading technology and best practice) are specific, sticky, tacit or leaky ("non-tradable"), these should by exploited by internal rather than by external service providers (i.e. by MNO' own service providers rather than by MVNOs) due to the extra transaction costs involved in dealing with external rather than internal service providers[101]. Only under the opposite conditions would we expect MVNOs to develop into a competitive alternative.

101 Uniquely valuable assets correspond to the essential facility concept in the telecom regulation literature. The latter are facilities that (i) are controlled by a monopolist, (ii) are considered a necessary input for the provision of downstream services, and (iii) cannot be duplicated in any technically or economically feasible way. According to the received doctrine, appropriate candidates for regulation are only those facilities that are "essential" in the above sense because of the power abuse potential (and associated super-normal profit potential) such facilities represent.

Figure 21. The Competitiveness Issue



MOBILE VIRTUAL NETWORK OPERATORS[102]

Only MNOs are licensed to run networks using the radio spectrum. The scope of mobile virtual network operators depends therefore on the number of additional elements of the mobile network that are outsourced to MNOs. This would range from the minimum use of the mobile networks' facilities to almost total dependence on them in which case a MVNO from a consumer point of view would scarcely be distinguishable from MNO's own service providers. Real economic dependence, however, will only occur to the degree alternative MNOs are non-existent (monopoly) or to the degree switching between them would be prohibitively costly (due to heavy investment in non-redeployable assets). Only under these conditions would extra safeguarding mechanisms be needed beyond those provided by simple market contracting.

At one extreme, all the MVNO will require is the use of the radio elements of the mobile networks and such fixed parts of the networks as are necessary to route calls between the radio elements and the first convenient points at which calls can leave (or join) the mobile networks on its way to (or from) the

102 The technical information source for this section is Oftel (1999a,b).

facilities of the MVNOs. This minimum amounts to no more than, firstly, the radio transmission link, its control functions and the mobility management functions that keep track of exactly where mobile handsets are located so that calls can be delivered to them; and secondly, some transmission and switching facilities needed to link the radio facilities to the points of interconnection either with the MVNOs' systems direct, or with transit carriers. This would maximize the MVNO's control over its customers and their calls.

At the other extreme, calls from MVNOs' customers will be handled virtually entirely by the mobile networks as if in fact they were calls from subscribers to the mobile networks. This would imply maximum use of the mobile operators' facilities and minimum investments by the MVNO. In particular all the verification operations would be carried out by the mobile operators whose databases would have to be geared to receive, process and supply data concerning the MVNOs' customers as well as their own. In the same way, the routings used for the transport and delivery of the calls after leaving the mobile networks to reach the terminating operators would be exactly those chosen for the calls originated by the mobile networks' own customers. Virtually all that would be different is that the billing and service performance information would have to be transferred from the MNO to the MVNO. In fact, all the MVNO would directly control would be the SIM card which contain the information that identifies the user to the network, especially the Mobile Network Code (MNC) that identify the network issuing the SIM card, besides some mission-critical databases such as their own Home Location Register (HLR) and Authentication Centre (AUC) functions.

From a transaction cost economics point of view, however, the critical factor is not the share of facilities and services outsourced to external network operators. The critical factor is rather the MVNO's own investment in MNO-specific (non-redeployable) applications and capabilities that would lock the MVNO to his chosen MNO for a substantial period, eliminate seamless network switching and create potential hold-up situations at recontracting intervals. Whereas network facilities and service applications are the hardware and software elements needed to produce and deliver telecom services to final users, technical interfaces are those elements that interconnect (more or less seamlessly) the hardware and software of the MVNO with those of the MNO or other complementary network operators. In this respect, a full and unconditional opening of the value-chain essentially means turning proprietary and closed interfaces into non-proprietary and open ones, while simultaneously offering network facilities and service applications that newcomers can afford to rent or buy. Under exclusive contracts, the value-chain will normally be closed to all except the selected

partner, whereas the underlying core technology with associated interfaces may still be more or less proprietary (Spiller 2001)[103].

Most relevant early examples of less standardized technologies were technology enablers such as SIM Toolkit and CAMEL that would enable MVNO subscribers to maintain the functionality of their handsets and their access to value-added services when visiting different host networks (MNOs)[104]. For example, the SIM Toolkit standard provides only a technical standard for communication between SIM card and handset, not for communication between SIM card and the mobile network (for downloading and service upgrading), nor for the development of the respective service applications. Consequently, closer collaboration will be needed for the service transactions between regular and virtual operators than between regular operator and simple resellers. As value-added services grow in number, complexity and importance, the MVNO will be increasingly closer tied to his MNO counterpart across proprietary and operator-specific (less open) interface standards. Besides, exchange of an increasing number of complementary, interdependent and interactive services among partially competing players will complicate pricing issues and performance monitoring and increase transaction hazards even further.

To profit from virtual network operation, the price margin defined by the difference between final user price and network rental price must cover not only added transaction costs, but also the MVNO's own production cost[105]. In this case, outsourcing and specialization may contribute both to production cost savings and to revenue growth. Through outsourcing of network elements and service applications costly duplication can be avoided, existing infrastructure more fully exploited, and network services provided at lowest possible unit cost (economies of scale). MVNOs may then gain extra profit if cheap transport and access services are combined with advanced functionality and value-added services, produced by the virtual operator's

---

103 Except for the renting and pricing issue, opening the value-chain in telecoms is strikingly similar to opening the value-chain in the computer industry. In particular, as telecoms converge with computing, lessons from the computer industry may increasingly become relevant for the telecom industry.

104 According the definitions used by Oftel (1999b): SIM (Subscriber Identification Module) – a small smart card device which is fitted into the mobile handset (usually on purchase), and is required in order to allow the user to make normal calls. SIM Toolkit – new GSM SIM card capability that allows operators to use the full computer processing power of a SIM card to create new services. CAMEL (Customised Applications for Mobile Network Enhanced Logic) – allows roamed customers to access the value added services of their home network, even if the visited network does not support those services.

105 Oftel (1999b) suggests Retail-Minus - a charge based on the retail price to the end user minus the costs of all elements of the call which are no longer supplied by the mobile network operator (since they will now be supplied by the MVNO).

own facilities and capabilities. The frequently acclaimed synergy effects from consolidation may even turn negative to the degree virtual operators develop into a more efficient solution for new service creation and network utilization.

Inter-firm specialization may provide additional benefits. MVNOs that rent, rather than own basic infrastructure, will be compelled to specialize within a narrower field of activities such as within intelligent facilities and more advanced service applications. Profit margins will generally also be higher from operating their own intelligent facilities and advanced applications than from reselling network services produced by external operators. However, specialization from virtual operation will not only create positive learning and income effects, but also increasing transaction costs when opportunistic and interdependent players interact across non-standard interfaces. Still, many virtual operators may connect with regular radio-based operators across rather complex and non-standardized interfaces. The more complex and the less standardized the interfaces, the higher the transaction hazards both due to more complicated pricing problems and to more consequential dependency relations[106]. Even with lower roaming prices (e.g., closer to average long run incremental cost), increasing competition and transaction costs may eradicate most of the MVNO's profit potential.

Summarized, both high roaming prices and high transaction costs may prevent MVNOs from becoming profitable. To improve their prospects, sources of transaction costs should be eliminated and contractual and regulatory safeguards enforced that prevent transaction costs from escalating and market power from being extended and abused. Some incumbents have even started to explore the conditions for MVNO-profitability more extensively, particularly in foreign markets where the roles are reversed and the need for an alternative MVNO-strategy is imperative[107].

EARLY DEVELOPMENT

The MVNO concept was pioneered by MCI in the US in the 90 and copied by Sense and Virgin Mobile in Europe a few years after. Then as now, to profit

106 *To create a virtual network operator, access not only to the incumbent's physical network but also to his Operational Support Systems (OSS) is required. These are computer databases and systems that provide services and network management, administration, planning and repair functions, as well as functions related to customer operations, such as customer care and billing.*

107 *This led Telenor and BT to develop a business plan for a software-based telco, named Facet, which illustrate perfectly the MVNO concept (Halbo & al., 1999a,b). At that time, Telenor also ran its own independent (semi-virtual) mobile service provider, named Zalto, focusing on the youth segment.*

from virtual operations, MVNO's retail prices had to cover MNO's rental prices, MVNO's production costs plus any additional transaction costs associated with renting the MNO's facilities. Since retail prices in the start up phase did not cover total production costs for smaller volumes (less than 200.000 average voice subscribers), longer term profitability would essentially depend on creative marketing and innovative value-added services that would increase revenue, and on standard interfaces and appropriate contract schemes that would restrict transaction costs. Achieving both higher revenue and lower transaction costs under simple market contracts proved difficult, however, since increasing sales of innovative value-added services did not only increase revenue, but also required more complex and less standardized interfaces that increased transaction costs.

Plausible solutions were either to specialize on "commodity" voice and data services transmitted across standard interfaces at minimal transaction costs, or to specialize on innovative value-added services transmitted across non-standard proprietary interfaces supported by long term inter-firm contracts that prevented transaction costs from escalating. Under more protective and mutually rewarding contracts, virtual operators might develop into a more efficient and creative marketing force than what vertically integrated units of radio-based operators could achieve, thus generating extra profit not only for themselves, but also for their radio-based partner. Subsequently, this might enable MVNOs to develop competitive advantage in value-added service, specialized branding and marketing, and MNOs to develop competitive advantage in basic mobile network operations.

RECENT DEVELOPMENT

Recent industry developments provide at least some support to the above hypotheses. After being pioneered by CMI and copied by Sense and Virgin, the MVNO practices are now rather widespread, with a surge in the number of launches in the past three years. According to research by Blycroft (2007) there are now more then 250+ active MVNOs in the world. Of these 150+ are in Europe, 60+ in North America, and 40+ in the rest of the world (Australia, Hong Kong, Israel, Japan, Malaysia, New Zealand, Réunion, Russia, Singapore, South Africa, Taiwan, Ukraine, Zanzibar)[108]. Increasingly,

108 *As pointed out by Buckland (2007): "Until recently, most MVNOs were similar and offered simple, no-frills services, often based on prepaid voice at prices undercutting the incumbents' offerings. A number of companies made a success of this business model and there is still scope for it to be used efficiently in some markets, especially by organisations such as large retail groups with well-known brand names. However, some new MVNOs have shunned the no-frills business model altogether and are concentrating on offering data and content services to niche markets. At the same time, fixed and cable operators are increasingly*

opening of the value-chain has also spread beyond virtual network operations into complementary mobile devices, applications and value-added services, thus stimulating further MVNO launches, as illustrated in the US mobile media and entertainment (MME) market. Verizon Wireless' recent announcement that it will begin accepting third-party applications and non-Verizon-branded devices on its CDMA network from late 2008 is apposite[109]. These and similar announcements and initiatives by Verizon led Rehak (2007) to conclude that "*Verizon has recognized that carriers can no longer compete effectively in the US market with a 'closed-shop' service and device provision model. Verizon's decision to open up the network to new devices and applications is a positive development, and we believe that other carriers in the US market will launch similar initiatives sooner rather than later."* Notably, to speed up the development of the MME market, Verizon and similar US carriers are increasingly relying on the creative services of mobile virtual operators enabled and stimulated by the carriers' own specialized MVNO programs (e.g.; Verizon Wireless' Agent and MVNO Programs). However, to survive in increasingly more crowded MVNO markets, new entrants have to offer unique propositions.

---

choosing to extend their service portfolios by using the MVNO model to enter the mobile market."

109 *As pointed out by Alexandra Rehak, US Research Director of Analysys, the global advisers on telecoms, IT and media (Rehak 2007): "Verizon's decision enables the carrier to focus on its strengths by transferring some of the development and marketing burden to its third-party partners in exchange for fairly limited investment on the carrier's part. Verizon will also be able to maximize its ARPU and network usage by opening up to new mobile data devices – particularly as the network is upgraded to 3G LTE. More critically, at least with regard to short-to-medium term ARPU growth, Verizon can use the new policy to help drive development of off-deck mobile media services, which are an essential part of creating an attractive mobile media and entertainment (MME) proposition. The top-selling content will probably always be from the top of the carrier deck, but customers need to discover content in other ways if the MME market is to achieve its full potential in the US. Off-deck providers can experiment and specialize in ways that would not be feasible or economical for carriers. Carriers will ultimately benefit by creating more opportunities for off-deck MME offerings to succeed, rather than short-sightedly focusing on maximizing their immediate revenue share".*

CONCLUSION

This paper have analyzed and exemplified mechanisms and conditions that may turn virtual network operators into profitable businesses. To profit from virtual operations, MVNOs must either perform their specialized "virtual network operating" functions more cost efficiently or provide content service (e.g.; mobile media and entertainment services) of higher value than achievable by ordinary MNOs which, on their part, are more likely to profit from specializing on value-chain coordination and remaining network functions. Regulatory enforced mobile network access supported by interface standards and firm-like contractual safeguards were among the main factors that initially affected such MVNO profitability. As regulation and technical interfaces develop into more generally acceptable and reliable standards, less tightly integrated relations may suffice. The recent global surge in MVNO launches offering more content rich mobile media and entertainment services provides illustrative and supportive evidence.

REFERENCES

Barney J.B. (1997), "Gaining and Sustaining Competitive Advantage. Reading", MA: Addison-Wesley.

Blycroft (2007), "The MVNO Directory 2008" accessible via: www.mvnodirectory.com/.

Buckland A. (2007), "Strategies for MVNOs" London: Analysys Report (accessed March 10, at http://research.analysys.com/default.asp?mode=products.asp).

Chesbrough H.W, Teece D.J. (1996), "When is Virtual Virtuous", Harvard Business Review, January-February, 65-73.

Financial Times Survey, FT Telecoms (2000a), "Nextcall: Virtual Newcomer's Battle", by Richard Handford, June 21th.

Financial Times Survey, FT Telecoms (2000b), "Virgin Mobile: On a Fast Track to Ring up Profit. A Combination of Simple but Cheap Tariffs and Advanced Services Lies at the Heart of the UK Company's 'Virtual' Strategy", by Neil McCartney, June 21th.

Foros Ø., Hansen B. (2000), "Connecting Customers and Disconnecting Competitors. The Facility-Based Firms' Strategy towards Virtual Operators", Kjeller, Telenor R&D Notat N 33/2000.

Halbo G., Mallett C., Freestone M., Trill M. (1999a), "Exploiting the Edge in intelligence to build a SoftTelco business", Kjeller, Telenor R&D Report, R 108/99.

Halbo G., Mallett C., Freestone M.,Trill M. (1999b). "Business plan for 'Facet,' a Software based Telco", Kjeller, Telenor R&D Report, R 109/99.

Liebeskind J.P. (1995), "Knowledge, Strategy and the Theory of the Firm", Strategic Management Journal, Winter, Special Issue, 17, 93-122.

Oftel (1999a), "Mobile Virtual Network Operators: Oftel Inquiry into what MVNOs could offer Consumers", a Consultative Document issued by the Director General of Telecommunications, June.

Oftel (1999b), "Oftel Statement on Mobile Virtual Network Operators", October.

Rehak A. (2007), "Verizon's open-network policy reflects the new mobile market reality, Analysys Insight.

Teece D.J. (1986), "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy", Research Policy, 15, December, 285-305

Spiller P. T. (2001), "Wireless versus Wireline. Micro-Charging and the Evolution of the Internet", SNF Yearbook, 2001, 113-122. Bergen: Fagbokforlaget.

St.meld.nr. 24 (1999-2000), "Om tilgang til mobilnett og om innføring av tredje generasjons system for mobilkommunikasjon". (in English: "About access to mobile networks and about building third generation system for mobile communications").

Williamson O.E. (1991), "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives", Administrative Science Quarterly, 36, June, 269-296.

Williamson O.E. (1999a), "Strategy Research: Governance and Competence Perspectives", Strategic Management Journal, 20, 1087-1108.

Williamson O.E. (1999b), "Public and Private Bureaucracies: A Transaction Cost Economics Perspective", Journal of Law, Economics and Organization, 15, 306-342.

Mobile Telecommunications Access and Growth in Developing Countries: the Role of Regulatory Institutions[110]

FEDERICA MAIORANO
Centre for Competition and Regulatory Policy (CCRP), Department of Economics, City University, and Ernst & Young.

JON STERN
Centre for Competition and Regulatory Policy (CCRP), Department of Economics, City University.

Abstract: *This paper studies the relationship between regulation and performance in the mobile telecommunications sector, taking account of the economic impact of telecommunications infrastructure on aggregate income and of the role of country institutions in promoting economic growth. We address these questions by estimating a system of equations for a panel of 30 low and middle-income countries over the 1990 - 2004 period. In summary, the evidence we present confirms the positive effect of regulatory institutions on telecommunications penetration and also highlights the contribution of a more widespread mobile telecommunications infrastructure to higher levels of GDP per capita.*

INTRODUCTION

Mobile communications have enjoyed impressive rates of growth across low and middle-income countries in recent years. The average number of phones per 100 population in this sample has increased from less than 1 in 1990 to around 40 in 2004. Compared to the availability of fixed lines (around 17 lines per 100 population as of 2004), the success of mobile communications is even more staggering.

High connection charges and long waiting lists, as well as the substantial investment required to develop extensive fixed networks, have held back traditional communications networks and favoured the expansion of mobile services. In addition, mobile markets have often been characterised by a relatively competitive market structure almost from service launch, while the liberalization of fixed markets has somewhat lagged behind. The importance of competition among mobile networks is highlighted by research published by the World Bank[111], which found that the introduction of a second and

110 The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and they do not represent the view of the organizations to which the authors are affiliated.

111 Gebreab (2002) relies on a dataset of 41 African countries.

subsequent competitors accelerates mobile penetration, while the presence of a state-owned telecoms incumbent in the market constrains it.

However, the continuing success of mobile networks as an effective substitute to fixed services in the future is not entirely clear, as some factors may inhibit its further development. Firstly, while spectrum may not prove to be a scarce resource in the early days of network deployment, some countries (e.g. India) have begun to suffer from network congestion. Secondly, broadband services are currently based mostly on fixed networks (ADSL, cable or Fibre to the Premises). Although mobile voice and SMS text use will continue to grow at a strong rate in developing countries, it is, though, still unclear whether the penetration and usage of mobile data services will become established in these countries over the next 5 – 10 years. For instance, 3G licences have yet to be issued in China and India. However, as 3.5G is being positioned as a broadband service, it may have the potential to capitalize on low broadband penetration in emerging markets.

The key role of mobile communications lies in the transformations it brings to everyday life and making business, especially in the absence of reliable and widespread fixed telecommunications networks. In Africa, mobile phones are used to provide public phone access, particularly in rural areas. Moreover, their importance for improving information on market prices in different locations is documented by a variety of case studies. In a recent empirical paper, Jensen (2007) shows how the introduction of mobile phones has significantly improved market efficiency in southern India, e.g. in the fisheries sector. There is also much case study and anecdotal evidence to suggest that the same is true for a wide range of agricultural products.

While the evidence for the economic impact of mobile communications appears increasingly well-established, understanding the factors leading to higher penetration rates appears difficult – we observe countries with very different socio-economic characteristics having similar proportions of mobile subscribers. One possible explanation may be found in the wave of telecom sector reform which, in the mid-90s, led many countries to commercialise and (in many cases) privatise their national incumbent telecom company; to liberalise telecom markets; to introduce competition, particularly in the mobile sector; and to establish separate (non-Ministry) regulatory agencies.

Much attention has been devoted by international institutions, such as the World Bank, to the question of how to reform a sector so as to improve access to infrastructure, be it communications networks or energy and other utilities. Within this reform programme, the two most important issues are (a) the design of the market framework for the sector; and (b) the establishment of a separate regulator, i.e. one that is autonomous (or “independent”) *both* of the government *and* of the incumbent national telecom company.

In assessing the success of such telecom reform programmes, we need to be able to find out whether the presence of a separate regulator can be demonstrated to lead to higher access to infrastructure. In the case of mobile communications, this means assessing where a separate regulator leads to higher mobile penetration rates. This issue can be addressed in different ways, including case studies and purpose-designed policy audit. These types of study provide detailed and thorough information on specific countries or policies and tend to show a positive relationship (e.g. World Bank, 2005). However, in order to reach more general conclusions we rely on a formal econometric analysis of the underlying relationships between economic variables for a number of countries.

In what follows, we summarize the conclusions of an econometric analysis of the interrelations between regulatory institutions, mobile penetration and GDP levels in a sample of 30 middle and low income countries over the period 1990-2004[112]. We also very briefly refer to some case study results, which highlight the importance of regulatory institutions in infrastructure sectors.

The relationship between the existence of a separate regulator and infrastructure industry investment and productivity levels has been a particular focus of debate. This issue has been investigated in numerous papers, with reference both to the telecommunications and to the electricity sectors (See, for instance, Gutierrez 2003; Cubbin and Stern 2006). These studies focus on the characteristics of regulatory institutions that one would expect to be associated with higher performance on key outcomes such as access levels, investment, efficiency and growth. For telecoms, a key outcome is the number of telephone lines per 100 inhabitants – fixed and/or mobile. However, this literature does not, with some exceptions, pay much explicit attention to the institutional setting within which the new regulatory agencies operate such as political structure, the rule of law or the degree of economic openness. We look at this both directly (i.e. with some 'country governance' variables) and indirectly (via the use of country specific fixed effects) which pick up the impact of unmeasured factors, including country-specific institutional quality.

The paper takes a fresh look at the relationship between regulation and performance in the telecommunications sector and, in particular, to the role of the institutional setting. In consequence, we take seriously the argument that the economic impact of telecommunications infrastructure on the growth of aggregate income is closely interrelated with the role of country institutions in promoting economic growth.

112 For details of the estimated model, see Maiorano & Stern (2007).

Our main conclusion is that our analysis provides strong additional evidence for the claim that a separate regulator significantly helps increase telecom penetration rates in developing and transition countries, in this case for mobile communications. Our study also provides evidence which strongly supports the claim that higher levels of mobile penetration increase GDP per capita.

In what follows, we summarise the approach adopted in the paper in Section 2, discuss the main results in Section 3 and make a few short concluding comments in Section 4.

MAIN ISSUES AND METHODOLOGY

The present work studies the penetration of telecommunications infrastructure, as measured by *the number of mobile subscribers per 100* population in 30 developing countries in Africa, Asia, Latin America and the Caribbean, and Central Europe.

Among the factors that may affect penetration, we consider the effect that income may have on the uptake of mobile telephone subscriptions. This is related to the question of whether penetration is demand-constrained or supply-constrained in developing countries. Case-study evidence, as well as the substantial waiting lists for fixed telephone lines, points to supply-side limits, rather than low levels of demand, as the main reason for the limited penetration of telecommunications services in developing countries. Hence, introducing an effective regulator should help encourage investment which will alleviate the supply bottlenecks.

Another consideration relates to the possible feedback effects between penetration and income. Higher income levels may increase the demand for mobile services. In addition, the expansion of network infrastructure services, including telecommunications, plays a crucial role in economic development[113]. There is now considerable evidence that higher telecommunications capacity – fixed and mobile – can have sizeable effects on the level and rate of growth of national income[114].

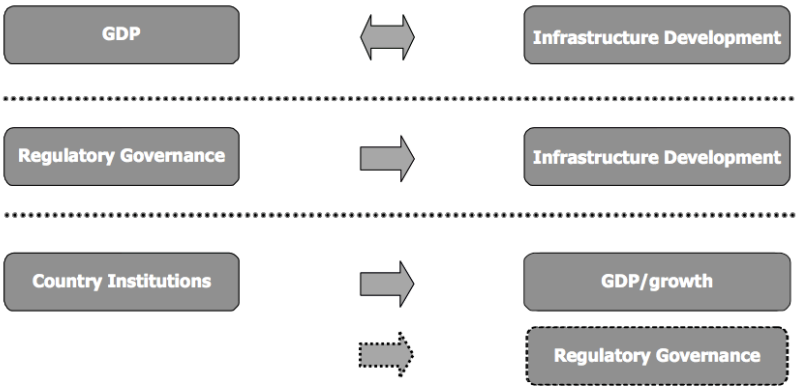
The importance of the telecommunications sector in improving a country's income level is a major complication when analyzing the factors that influence telecommunications penetration rates. While it is conventional to assume that income is among the variables affecting the demand for

113 See Canning (1999) and Canning & Bennathan (2000).

114 See Correa (2006), Esfahani & Ramirez (2003), Röller & Waverman (2001) and Waverman & al. (2005).

infrastructure capacity and services, the economic feedback impact of telecommunications infrastructure capacity also needs to be modelled if we are not to have a misleading picture. This is represented by the two-sided arrow in the top row of the figure below.

Figure 22. Factors Potentially Affecting Infrastructure Development



Considering the second row, the economic importance of the telecommunications industry has been among the factors contributing to the active role of governments in this sector. The reform process that has taken place in developed countries and in many low and middle-income countries aims at achieving public interest targets by a set of major policy changes, in which the establishment of a regulatory framework is accompanied by sector restructuring, the liberalization of the market and the privatization of the incumbent. Put at its simplest, introducing private finance and privatizing expanding telecommunications industries has been the main force behind the development of new regulatory organizations as well as, arguably, encouraging general improvements in country governance in the areas of commercial law enforcement.

Our study concentrates on the regulatory framework for telecommunications per se and, in particular, on key aspects of regulatory governance (e.g. the existence of a separate regulator). Regulatory reform has the stated objective of promoting better infrastructure development, among other targets, by attracting investment and lowering the cost of capital. This effect is symbolized by the arrow in the middle panel[115].

115 It may be argued that countries with more widespread telecommunications penetration are more likely to set up regulators and that therefore there may be

Effective regulatory frameworks need to be adapted to the specific circumstances of each country and, in particular, to their governance characteristics. This is shown in the third part of the previous. One interpretation of this relationship views the establishment of an independent regulator as a substitute for strong country institutions (e.g. strong property rights and competition authorities, supported by strong and independent commercial courts). However, an alternative and perhaps more plausible view is that countries with strong institutions may be more likely to engage in substantive reform, which will include genuinely independent and high quality sectoral regulatory agencies. This last relationship is represented by the dotted lines in the lower panel of the previous figure.

Finally, as highlighted by widespread evidence of the positive impact of high-quality country governance and institutions on GDP growth rates, we need to include the potential direct effect of country institutions on income into the analysis.

The present study attempts to bring together these questions into a unified framework of analysis. We do this by estimating a system of equations including all of these factors for a panel of 30 low and middle-income countries over the 1990 - 2004 period.

SUMMARY OF THE APPROACH

In order to deal with the interactions described above and represented in Figure 22, a system of simultaneous equations is estimated in which the dependent variables (i.e. the variables whose time-path we try to explain) are:

- Infrastructure development – mobile penetration rates;
- Income – GDP per capita; and
- Regulatory governance – whether or not countries adopt a separate regulator .

Our approach assumes that these variables are endogenous, i.e. they are assumed to be determined *within* the model, as shown in Figure 22 above, rather than as being given from outside it.

The basic specification consists of three equations, one for each of the three variables above. We estimate the equations both separately and also jointly

some feedback effects from infrastructure development to regulatory governance. In fact, this apparent feedback may instead be related to other factors, such as liberalization or privatization, affecting both regulatory governance and infrastructure development.

as a system. The system approach allows us to take full account of the feedbacks between the endogenous variables shown in Figure 22.

In the first equation, mobile penetration rates in each country are explained by income (GDP per capita), regulatory governance and other variables. Regulatory governance, in the remainder of the paper, is proxied by a dummy variable indicating the establishment of a regulator as a separate entity from the Ministry in charge of the sector[116]. We also tried some other telecom regulatory indicators, but the 'separate entity' dummy variable was the most successful[117]. The dummy takes the value of 1 for any country that has a separate regulator in the years it was present (e.g. Kenya 1999-2004) and zero otherwise.

Other variables that we include to explain mobile penetration rates include, firstly, the price of mobile services and of fixed services (in both cases, average revenue per subscriber); and, secondly, dummies for majority privatization of the incumbent and the liberalization of fixed long-distance services.

We also include in all equations dummy variables for each country (a so-called "fixed effects" approach) in order to control for unobserved characteristics and omitted variables that are specific to a given country. Previous work has shown that these country-specific fixed effects are important in providing a good explanation of telecom access rates and the role of regulatory institutions – a result we also find. Indeed, much of the effect of variations in the quality of country governance probably comes through via *indirect* effect of the individual country-specific fixed effects. Finally, we include year dummies to take account of time effects that are common across countries.

The variables regarding sector reform are summarized in Figure 22 below. This simple comparison shows higher mobile penetration in countries that have implemented different types of reforms (telecommunications law, separate regulator, liberalization of long-distance services and majority privatization of the incumbent) compared to the others. At first glance, this

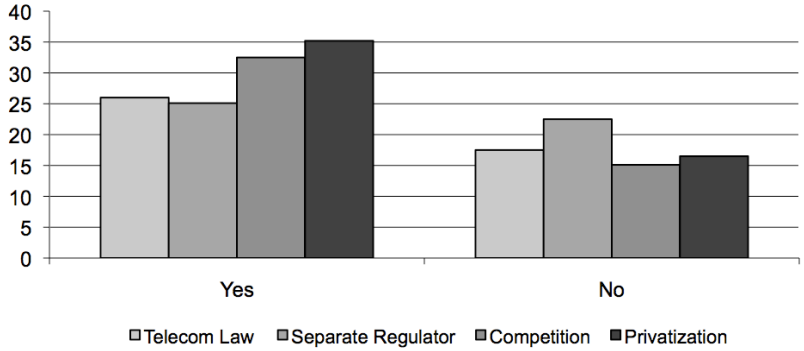
---

116 This variable should not be confused with what is reported in the ITU Development Database as "autonomous decision-making". While autonomy involves a subjective judgement on the functioning of the regulator, our variable merely reflect the existence of a separate entity.

117 These are: whether (a) the country has passed a framework law for the telecommunications sector; (b) the country has established a regulator as a separate entity from the policy maker; and (c) the regulator is not funded by the Government's budget. In addition, we also tried using the number of years since the creation of the regulator to capture the time necessary to build up staff numbers and competences and reputation.

would suggest these types of reform may well have had a positive impact on penetration and we test this in our estimation.

Figure 23. Mobile subscribers per 100 inhabitants



Source: ITU, regulator's websites

The second equation relates income levels (GDP per head) to the penetration of mobile communications, a measure of country institutions[118] and other variables. The institutional characteristics considered in this study are: (a) protection against expropriation risk (e.g. an index of constraints to the executive; see Henisz 2002); (b) financial market development (e.g. share of credit to the private sector on GDP). The key point is to test not only whether separate regulation increases mobile penetration rates but also whether higher mobile penetration rates increase income levels.

Finally, in the third equation, we try to explain regulatory governance (the adoption by countries of a separate regulator) by per capita income, whether or not the incumbent telecom company has been majority privatised, country institutions and some other variables. The other variables that we consider include pressure by international organizations (e.g. conditionality conditions imposed by international financial institutions), as proxied by multilateral lending.

118 The endogeneity of institutions is a hotly debated topic. In the model that we estimate, it is assumed that country institutions are pre-determined. For this reason, we include in the system the lagged variable, which pre-date the period of analysis, rather than the current level on the grounds that institutions in previous years cannot be affected by income levels in subsequent periods (see Rajan & Zingales 1998; Esfahani & Ramirez, 2003 for previous use of this approach).

KEY RESULTS

In this Section, we summarize and interpret the results of our analysis.

In the first equation, we estimate the impact on mobile penetration rates by income (GDP per capita), regulatory governance (separate regulator) and other variables. We find some evidence that the existence of a separate regulator increases penetration rates for mobile telecommunications in developing countries, although the precise estimate of the effect varies considerably, depending on the exact specification.

Depending on the specification, we estimate that having a "separate regulator" increases mobile penetration rates of between 44% to around 171%. On a conservative basis, in our sample of developing and transition countries, the presence of a separate regulator is associated with mobile penetration about 45% higher than in comparable countries without a separate regulator.

Our results are less robust than in previous studies of the impact of regulators on fixed line penetration rates. This variability may be due to our relatively simplistic regulatory variable or to the possibility that the role of regulators is not as crucial for mobile operators as it is in the fixed market[119].

In the first equation, we also obtain estimates for the coefficient on GDP per capita. However, these are not statistically significant in a fixed effects model. This finding is in line with the observation that mobile penetration has boomed in poor countries even at times when income growth was negligible.

We also find that mobile penetration rates are higher with lower mobile prices (as proxied by average revenue per subscriber) and also with higher telecom investment rates. The existence of a separate regulator is likely to encourage higher investment and may well help keep prices lower where competition is limited, so these are other channels by which a separate regulator may have a positive effect on mobile access in developing countries.

The negative coefficient on the dummy 'liberalization', which indicates the opening to competition of long-distance services provided on fixed networks, seems to suggest substitutability between fixed and mobile services, which is particularly likely in developing countries. However, this result is not robust

119 In fact, in a large panel of countries, Estache & al. (2006) find that the dummy for a separate regulator does not have a significant effect on fixed telecommunications penetration.

as it does not hold when we estimate the model on a larger sample of low and middle-income countries[120].

In the second equation, income levels are explained by the penetration of mobile communications, a measure of country institutions and other variables. We find a sizeable and strongly significant impact of mobile telecoms infrastructure on per capita GDP in our sample. This result confirms the findings of other economists on the economic impact of mobile telecoms in developing countries (e.g. Waverman & al. 2005).

Depending on the specific model, we estimate a coefficient on GDP per capita on mobile penetration which ranges from 1.3% to 4.4%. By combining this finding with the results for the first equation, *we can conclude that, on average in our sample, the establishment of a separate telecoms regulator and the effect of that on mobile penetration is associated with GDP per capita that is around 0.5% higher than it would have been otherwise*

In the third equation the establishment of a separate regulator is explained by income, country institutions and other variables. The only significant variables are liberalization of fixed services and the privatization of the incumbent, which as expected are associated with higher probability that the country establishes a separate regulator.

Surprisingly, we find no evidence of any systematic link between country institutions and regulatory governance, but our country governance data were rather less good than most of the other data we used. In addition, country governance variables are usually very persistent over time and, in our model, their effect may well have been captured by the country specific fixed effects included in the equations[121].

Finally, we would like to add a word of caution as to the interpretation of our results. Even though we tried to take account of feedback effects, we cannot interpret the establishment of a separate regulator as the clear-cut *cause* of higher mobile penetration. This is a common issue in empirical economics and especially in the literature on institutions. Moreover, the results hold on average, i.e. in a small interval around the sample averages, and extrapolating the conclusions to countries not in the sample needs to be

120 *In a paper estimating demand for telecommunications services in developing countries, Garbacz & Thomson (2007) find that mobile services are not substitutes for fixed lines, but can be considered as complements. The effects of liberalization in the fixed market on mobile penetration is insignificant in a sample of developing countries analysed in Gasmi et al. (2006), who also find a positive effect of mobile competition on mobile access rates.*

121 *A recent study by Gasmi, Noumba & Virto (2006) has rather better data on country governance and finds significant effects of some country governance variables on mobile penetration rates and prices in developing countries.*

done with care, particularly since the results depend on unobservable country specific fixed effects. However, we have tested carefully for internal consistency and our results are also very similar to other econometric studies of the impact of telecom and electricity regulation in developing countries.

CONCLUDING COMMENTS

We have shown in the previous section that the existence of a separate telecom regulator in developing and transition countries is clearly associated with higher mobile penetration rates – around 45% higher than without a separate regulator. We also find that this induced higher mobile penetration is consistent with around 0.5% higher GDP per capita.

Our results are also consistent with the results from in-depth case studies. For instance, the impact of regulation on telecom development has been studied in Jamaica, in a number of countries in Africa and elsewhere and the results on the role of regulation are of the kind that one would expect from our econometric results. (See Shirley & al. (2002) for the role of regulation in telecom reform in Uganda and NERA (2004) on Mauritania and South Africa).

The case studies also investigate the role of regulators in introducing competition in fixed and mobile services and whether or not regulators provide a clear framework for investment. These are issues that arise within our econometric model but which we were unable to pursue in depth. However, this shows how the econometric work can complement case studies and the results of ex post regulatory evaluations[122]. For telecoms in general and mobile in particular, they are providing similar conclusions – which gives a lot more confidence in the underlying policy message.

In consequence, we conclude that having a separate regulator clearly seems to result in higher mobile access rates in developing countries.

On the role of competition, in some specifications we find some substitution between fixed and mobile services for developing countries, as the liberalization of long-distance fixed networks is associated with lower mobile penetration[123]. In addition, other papers provide evidence that more

122 See World Bank "Handbook for Evaluating Infrastructure Regulatory Systems" (2006) by Brown, Stern & Tenenbaum. This approach has since been applied to the evaluation of the Jamaican infrastructure regulator, including for telecoms and the development of mobile.

123 This result is not confirmed when a wider sample of developing countries is used.

competition between different mobile operators increases penetration rates (Gebreab 2002). However, for fixed markets, there is as yet relatively little competition – particularly infrastructure-based competition - in developing countries. Hence, it is difficult at this stage to draw conclusions as to the impact of competition between fixed and mobile networks in this context[124].

Policy makers in developing countries and their counterparts also need to consider wider effects of the growth of mobile relative to fixed services. A key issue here is the penetration and usage rates of Internet services, especially broadband. Mobile broadband is both expensive and likely to take some years to roll out in most developing countries.

The rapid growth of mobile usage in developing countries has yielded significant benefits to small producers as well as to households. It remains to be seen how this process will be taken forward. In particular, it is unclear whether the next stages of telecom (fixed and mobile) usage will be based on top-down roll-out of new services by existing operators. Alternatively, it is possible that more use will be made of a decentralised approach based on entry and innovation at regional and local level by new operators. The latter would imply a much larger role for facility based competition. If the latter model is adopted, the implications for regulatory design would be significant e.g. much more light-handed (and pro-competitive) regulation together with significant regional differentiation.

REFERENCES

Canning D. (1999), "Infrastructure's Contribution to Aggregate Output", World Bank Policy Research Working Paper No. 2246.

Canning D., Bennathan E. (2000), "The Social Rate of Return on Infrastructure Investment", World Bank Policy Research Working Paper No. 2390.

Correa L. (2006), "The economic impact of telecommunications diffusion on UK productivity growth", Information Economics and Policy, Vol. 18, pp. 385 – 404.

Cubbin J., Stern J. (2006), "The Impact of Regulatory Governance and Privatisation on Electricity Industry Generation Capacity in Developing Countries", The World Bank Economic Review, Vol. 20 No. 1, pp. 115 - 141.

124 See other papers in this volume for evidence on the impact of competition in EU and other developed countries in mobile and fixed services.

Esfahani H.S., Ramirez M.T. (2003), "Institutions, Infrastructure and Economic Growth", Journal of Development Economics, Vol. 70, No. 2, pp. 443-477.

Estache A., Goicoechea A., Manacorda M. (2006), "Telecommunications Performance, Reforms and Governance", World Bank Research Working Paper, 3822.

Gasmi F., Noumba P., Recuero Virto L. (2006), "Political Accountability and Regulatory Performance in Infrastructure Industries: An Empirical Analysis", World Bank Policy Research Working Paper No. 4101.

Garbacz C., Thomson H.G (2007), "Demand for Telecommunications Services in Developing Countries", Telecommunications Policy, 31, pp. 276 – 289.

Gebreab A. (2002), "Getting Connected: competition and diffusion in African mobile telecommunications markets", World Bank Policy Research Working Paper No. 2863. http://econ.worldbank.org/files/15963_wps2863.pdf.

Gual J., Trillas F. (2006), "Telecommunications Policies: Measurements and Determinants", Review of Network Economics, Vol. 2, Issue 2.

Gutierrez L.H. (2003), "The Effect of Endogenous Regulation on Telecommunications Expansion and Efficiency in Latin America", Journal of Regulatory Economics, 23, No. 3, pp. 257 - 286.

Henisz W.J. (2002), "The Institutional Environment for Infrastructure Investment", Industrial and Corporate Change, 11(2).

Jensen R. (2007), "The Digital Provide: Information (technology), Market Performance and Welfare in the South Indian Fisheries Sector", Quarterly Journal of Economics, Issue 3, pp. 879 - 924.

Maiorano F., Stern J. (2007), "Institutions and Infrastructure Development in Low and Middle-Income Countries: The Case of Mobile Communications", Utilities Policy, No. 15, pp. 165-181.

NERA (2004), "Framework for Evaluating the Effectiveness of Telecommunications Regulators in Sub-Saharan Africa", A Report for the World Bank (Global ICT Department).

Rajan R., Zingales L. (1998), "Financial Dependence and Growth", American Economic Review, Vol. 88, No. 3, pp. 559 - 586.

Röller L.H., Waverman L. (2001), "Telecommunications Infrastructure and Economic Development: A Simultaneous Approach", American Economic Review, Vol. 91, No. 4, pp. 909-923.

Shirley M.M., Tusubira F.F., Gebread F.A., Haggarty L. (2002), "Telecommunications Reform in Uganda", World Bank Policy Research Working Paper No. 2864.

Stern J., Trillas F. (2003), "Independence and Discretion in Telecommunications Regulation: Lessons from Independent Central Banks", Utilities Policy, Vol. 11, pp. 191-201.

Waverman L., Meschi M., Fuss M. (2005), "The Impact of Telecoms on Economic Growth in Developing Countries", The Vodafone Policy Paper Series, No. 3.

World Bank (2005), "Connecting Sub-Saharan Africa – A World Bank Group Strategy for Information and Communications Technology Development", Working Paper No. 51.

Infrastructure versus Service-Based Competition: the Case of Mobile Telecommunications

The right balance between competition via infrastructures or services is an ongoing and controversial debate among academics, as well as regulators. The exchange of ideas has induced innovative concepts, as the "ladder of investment" developed by Pr. Martin Cave. But, if discussions have led to useful guidelines in selecting regulatory schemes, fixed broadband has been primarily addressed up to now. This book is, therefore, a collection of essays assembled to highlight the specific dimensions of these issues within the field of mobile telecommunications. This industry is a singular example of "pure" competition via infrastructure, with very few other comparable cases, be it in telecoms, energy, railroad, postal services, air transport, … Thus, the issue in mobile telecommunications today is not that of creating progressive infrastructure competition, but that of introducing a degree of service competition … without deterring future investment in network infrastructures. European and non-European scholars have been invited to present their works to be widely circulated amongst regulatory authorities, as well as industrial players and the telecommunications research community. We hope that the studies presented in this book may shed light on future regulatory decisions and help to "fine tune" the regulation, so as to preserve the cycles of investment in the mobile networks infrastructures.

Quantifica Publishing - copyright 2008

ISBN : 978-2-9529662-1-4

Price : 20 €